doi:10.1006/jmbi.2001.5008 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2001) **312**, 927–934

JMB

COMMUNICATION

Energetics, Stability, and Prediction of Transmembrane Helices

Sajith Jayasinghe, Kalina Hristova and Stephen H. White*

Department of Physiology and Biophysics and the Program in Macromolecular Structure University of California at Irvine, Irvine, CA 92697-4560, USA

**Corresponding author*

We show that the peptide backbone of an α-helix places a severe thermodynamic constraint on transmembrane (TM) stability. Neglect of this constraint by commonly used hydrophobicity scales underlies the notorious uncertainty of TM helix prediction by sliding-window hydropathy plots of membrane protein (MP) amino acid sequences. We find that an experiment-based whole-residue hydropathy scale (WW scale), which includes the backbone constraint, identifies TM helices of membrane proteins with an accuracy greater than 99 %. Furthermore, it correctly predicts the minimum hydrophobicity required for stable single-helix TM insertion observed in *Escherichia coli*. In order to improve membrane protein topology prediction further, we introduce the augmented WW (aWW) scale, which accounts for the energetics of salt-bridge formation. An important issue for genomic analysis is the ability of the hydropathy plot method to distinguish membrane from soluble proteins. We find that the method falsely predicts 17 to 43 % of a set of soluble proteins to be MPs, depending upon the hydropathy scale used.

Keywords: membrane proteins; structure prediction; hydropathy plots; peptide bond; genomic analysis

Sliding-window hydrophobicity (HΦ) analysis[1] of the amino acid sequences of membrane proteins (MPs) is widely used to identify putative transmembrane (TM) helices. It is arguably the most successful and accurate protein structure-prediction algorithm developed in the past two decades. Nevertheless, TM helix identification by HΦ analysis remains naggingly uncertain.[2] It has long been recognized[3–5] that the results of HΦ analysis depend strongly on which of the many free energy (hydropathy) scales is used. Because assembled MPs are equilibrium structures[6–8] whose TM helices appear to be independently stable,[9] TM helix stability, and prediction, is fundamentally a problem of equilibrium thermodynamics. Solving this problem thus requires a complete accounting of the energetics of TM helix stability.

TM helix stability, defined as the free energy of transfer from the aqueous phase to membrane bilayer, is determined by the sum of the transfer free energies of side-chains and helix-backbone.[7,10,11] The importance of the backbone contribution was not clearly enunciated[12] until after the HΦ scales in common use were created. Although a recent theoretical estimate suggests that the dehydration of H-bonded peptide bonds is costly,[13] its exact cost is uncertain. In order to understand the consequences of ignoring this cost in HΦ analysis and to establish a value for it, we have analyzed the prediction accuracies of several hydrophobicity scales using a highly reliable MP topology database.[14] We find that the prediction inaccuracies of commonly used scales[3,15,16] arise from overprediction of TM helices, due to neglect of the unfavorable cost of dehydrating the helix backbone. Because this cost is accounted for in the Wimley-White (WW) experiment-based whole-residue hydropathy scale,[7,17,18] derived from measurements of host-guest peptide partitioning into *n*-octanol, the scale allows almost perfect identification of TM helices of MPs of known 3D structure. A novel analysis shows that the peptide bond dehydration free energy of the WW scale provides an excellent estimate of the helix-backbone membrane-insertion free energy. Because of the success of the WW scale, we expand it to include the ener-

Abbreviations used: TM, transmembrane; MP, membrane protein; HΦ, hydrophobicity; WW, Wimley-White; aWW, augmented WW; AA, amino acid.

E-mail address of the corresponding author: shwhite@uci.edu

0022-2836/01/050927–8 $35.00/0

getics of salt-bridge partitioning, which Wimley *et al.*[18] determined using the same family of host-guest peptides as for the WW scale. This augmented WW scale (aWW) will allow the consequences of salt-bridge formation to be considered in determinations of membrane protein topology.

Prediction accuracies of commonly used hydrophobicity scales

In order to examine the TM helix prediction accuracies of hydrophobicity scales, we created the MPtopo database and used a simple HΦ plot algorithm with a 19 amino acid (AA) window. MPtopo is a carefully curated database comprised only of MPs whose topologies have been validated experimentally.[14] The topologies of the 3D_helix set of MPtopo were determined crystallographically, whereas the topologies of its 1D_helix set were determined by gene fusions and other methods. A proper thermodynamic free energy scale should accurately identify stable TM helical segments in these two sets as those with favorable water-to-membrane transfer free energies ($\Delta G < 0$). Such segments should have a length that approximates the typical hydrocarbon core thickness of fluid lipid bilayers,[19] about 30 Å. Thus the 19 AA sliding window that is generally used in HΦ plot analysis, corresponding to a helix length of 28.5 Å (assuming a 1.5 Å rise per residue), is a reasonable choice for identifying TM segments.

Known transmembrane segments in the 3D_helix set of MPtopo have an average total length of 27(±6) AA. But determining the exact lengths of these TM segments was not an objective of our analysis. Rather, we simply searched the MP sequences for segments of 19 or more AAs with favorable water-to-bilayer transfer free energies, indicative of helical segments of sufficient length to span the bilayer. TM helix identification was considered successful if a known TM helix was identified irrespective of its exact length. Single-helix prediction accuracy was therefore used to quantify prediction success.[20] Exactly the same MPtopo data set, sliding-window length, and HΦ plot algorithm were used for all hydrophobicity scales (described in the legend of Figure 1). This assured an unbiased comparison, so that prediction accuracy depended solely on hydrophobicity scale. The formula used for prediction accuracy[20] accounts for incorrect predictions, as described in the legend of Figure 2.

Figure 1(a) shows that the three most commonly used scales, labeled EC,[15] KD,[16] and GES,[3] significantly overpredict TM helices in the 3D_helix and 1D_helix sets of MPtopo. Figure 1(b) discloses an inverse correlation between overprediction and prediction accuracy. In contrast, the WW scale[7] yields a close correspondence between the number of predicted and known TM helices, and consequently a high prediction accuracy. The most important feature of the WW scale, one distinguishing it from the other scales, is that it includes a cost for dehydrating peptide bonds. If this cost, greater than 1 kcal mol^{-1} per residue, is similar for TM helices, then side-chain-only scales (EC, KD, and GES) will overestimate TM helix stability by more than 20 kcal mol^{-1}. This will cause overprediction of the TM helices of MPs, as observed. Interestingly, Kyte and Doolittle[16] suggested that TM helices of MPs corresponded to HΦ plot peaks with a minimum (threshold) KD hydrophobicity of −1.6 kcal mol^{-1} per residue. Similarly, Engelman *et al.*[3] suggested a threshold of −1 kcal mol^{-1} per residue for the GES scale. These numbers can be rationalized by a backbone dehydration cost of ~1-2 kcal mol^{-1}, which is missing in these side-chain-only scales.

Estimating the thermodynamic cost of dehydrating the TM helix backbone

In the WW scale derived from the partitioning of small random-coil peptides into *n*-octanol, each whole-residue transfer free energy, ΔG_X^{WW}, is the sum of the glycyl backbone component ΔG_{glycyl}^{coil} (which is identical to ΔG_{Gly}^{WW}) and the side-chain component ΔG_X^{sc} (relative to Gly). The apparent success of the WW scale (Figure 1(a)) suggests that ΔG_{glycyl}^{coil}, 1.15(±0.11) kcal mol^{-1}, may be close to $\Delta G_{glycyl}^{helix}$, as suggested by White and Wimley.[7] In principle, the proposition that $\Delta G_{glycyl}^{coil} = \Delta G_{glycyl}^{helix}$ can be tested by direct measurements of the partitioning of model TM helices across lipid bilayers. But peptide solubility and aggregation[10] make this strategy problematic. We therefore adopted a different approach.

The known structures of MPs have delivered an unequivocal message about the importance of peptide-bond hydrogen bonding in MPs: TM segments can exist only as α-helices, β-barrels, or, in the case of gramicidin A,[21] β-helices. We reasoned that these structures might also tell us the best value of $\Delta G_{glycyl}^{helix}$, as follows. We carried out automated HΦ plot predictions of TM helices for the 3D_helix family of MPtopo, using as the hydrophobicity scale $\Delta G_X^{WW} + \Delta\Delta G_{COHN}$. The $\Delta\Delta G_{COHN}$ term, representing the difference between ΔG_{glycyl}^{coil} and $\Delta G_{glycyl}^{helix}$, was allowed to range from +0.4 to −0.4 kcal mol^{-1} in steps of 0.05 kcal mol^{-1}. Finally, we computed single-helix prediction accuracy Q as a function of $\Delta\Delta G_{CONH}$, assuming that an optimal value for $\Delta\Delta G_{CONH}$ would emerge.

Figure 2(a) shows that the number of helices predicted as a function of $\Delta\Delta G_{CONH}$ increases as $\Delta\Delta G_{CONH}$ decreases, demonstrating that underestimation of $\Delta G_{glycyl}^{helix}$ causes overprediction of TM helices. Figure 2(b) shows that Q passes through a maximum, consistent with the existence of an optimal value of $\Delta\Delta G_{CONH}$. Quite remarkably, the maximum Q is slightly greater than 99% and occurs in the vicinity $\Delta\Delta G_{CONH} = 0$. The most likely value of $\Delta G_{glycyl}^{helix}$ is therefore 1.15 kcal mol^{-1}, consistent with an insightful theoretical estimate.[13] To put this cost in perspective, ten leucine side-chains would be required to compensate for the

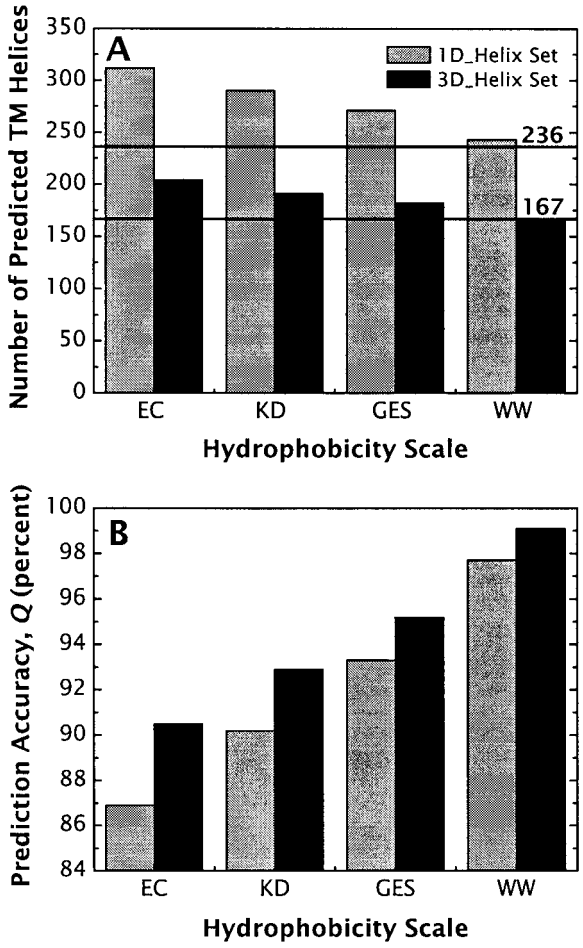

Figure 1. Summary of TM helix predictions of several hydrophobicity scales. Membrane proteins examined were from two sequence subsets of the MPtopo database (below), 3D_helix (black bars) and 1D_helix (gray bars). The results were obtained using automated hydropathy plot analysis (below) carried out with three commonly used hydrophobicity scales (EC, KD, and GES) and the WW whole-residue scale.[7] EC, Eisenberg Consensus scale;[15] KD, Kyte-Doolittle scale;[16] GES, Goldman-Engelman-Steitz scale.[3] (a) Total number (N_{pred}) of TM helices predicted by each scale. The horizontal lines indicate the known numbers (N_{known}) of helices. (b) Prediction accuracies (legend, Figure 2) of the four scales. The prediction accuracy of the WW scale slightly exceeds 99% for 3_D helix. MPtopo Membrane Protein Topology Database: MPtopo[14] database entries are divided into three subsets: 3D_helix, 1D_helix, and 3D_other. The sets 3D_helix and 1D_helix are comprised, respectively, of MPs of known 3-D structure, and MPs of unknown 3-D structure but known topology, as validated by various experimental methods. Being the most reliable data set, 3D_helix was used for computing the optimal value of $\Delta\Delta G_{CONH}$. The third set, 3D_other, consists of monotopic and β-barrel MPs of known 3-D structure, but it was not used in the work presented here. The most current database can be queried at http://blanco.biomol.uci.edu/mptopo. HΦ Plot Algorithm: Several HΦ plot algorithms of varying complexity were developed, but we settled on the simplest one because it tended to slightly under-predict TM helices while achieving maximum accuracy for all scales. Unlike the commonly used scales, the WW scale includes values for both the charged and neutral forms of Asp, Glu, and His. For it, assuming pH =7, we used hydrophobicity values for the charged forms of Asp and Glu and the neutral form of His. The sliding window of the HΦ plots was rectangular, with a width of 19 AA. (The detection algorithm performed optimally for this window length; shorter or longer windows reduced prediction accuracy for all scales.) The decision level[5] for TM helix selection for all scales was taken as $\Delta G = 0$, i.e. a favorable "peak" in a HΦ plot had $\Delta G < 0$. This is a strict criterion, and an essential one for an absolute thermodynamic scale, that has not been generally adopted for the commonly used scales. Major peaks in HΦ plots are usually comprised of several narrower noise-like peaks of varying amplitudes. These noise peaks must be accounted for by the peak-selection algorithm in order to identify the most likely position of the center of a putative TM helix. The criterion for detecting multiple narrow peaks was the existence of peaks within a half-window (nine AAs) of a chosen peak. The algorithm we used for TM segment identification is as follows: (1) identify all minima in the HΦ plot, (2) retain only those minima that have $\Delta G < 0$, (3) for minima that occur within nine AAs of other minima, retain only the one of greatest magnitude, (4) mark a segment of 19 residues centered on the minima retained in step (3), (5) combine any overlapping segments from step 4 to form one continuous segment, and (6) report segments identified in step (5).

23 kcal mol^{-1} backbone dehydration penalty associated with a 20 AA helix.

A similar procedure can be used to estimate the missing per-residue backbone contributions in the side-chain-only EC, KD, and GES scales. We found that optimal prediction accuracies ($Q = 98\%$) for these scales could be achieved using per-residue backbone dehydration costs of 1 kcal mol^{-1} for KD and GES, and 0.4 kcal mol^{-1} for EC. These results confirm the importance of the backbone in TM helix stability. However, we do not recommend the use of these "tuned" scales because the free energy values are taken from diverse thermodynamic measurements, and include various modifications of selected free energies based upon guesses and hunches about the physical behavior of side-chains in membranes. The WW scale, on the other hand, uses direct thermodynamic data derived from a set of consonant measurements.

Predicting biological TM helix stability

Even though the WW scale predicts TM helices accurately, can it predict the stability of single TM helices in biological membranes? For *Escherichia coli*, apparently it can. Because $\Delta G^{WW}_{Ala} = +0.5$ kcal mol^{-1}, polyalanine should not form a stable TM helix. But polyleucine should, because $\Delta G^{WW}_{Leu} = -1.25$ kcal mol^{-1}. The minimum hydrophobicity required for TM stability should therefore occur for some particular mixture of Leu and Ala residues. Chen and Kendall[22] found for *E. coli* that a 21-residue polyalanine segment could not

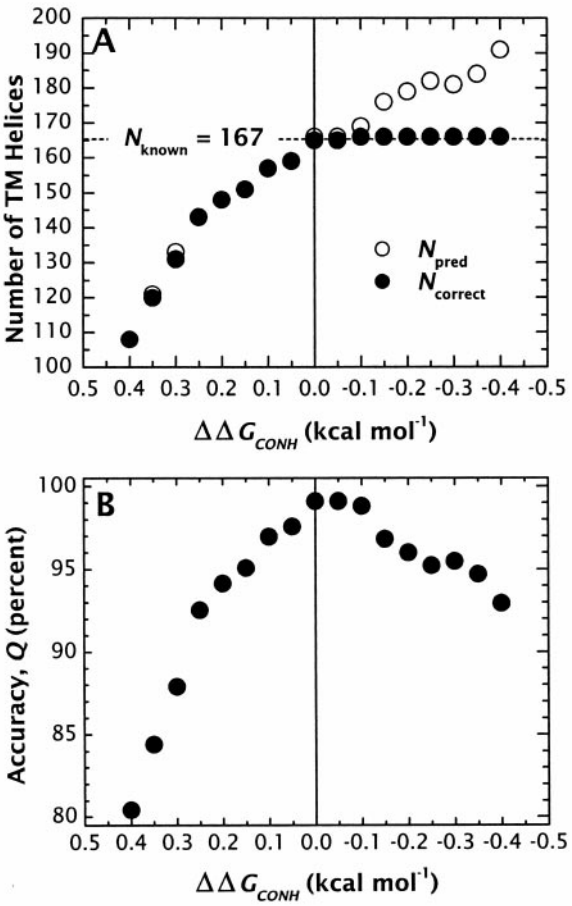

Figure 2. Optimization of the dehydration free energy of a helical glycyl unit ($\Delta G^{helix}_{glycyl}$) using the 3D_helix set of MPtopo.[14] $\Delta G^{helix}_{glycyl}$ was taken as $\Delta G^{coil}_{glycyl} + \Delta\Delta G_{COHN}$, where $\Delta G^{coil}_{glycyl} = 1.15$ kcal mol^{-1}. For values of $\Delta\Delta G_{CONH}$ ranging from -0.4 to $+0.4$ kcal mol^{-1}, TM helices were identified using an automated hydropathy plot procedure. (a) Number of predicted (N_{pred}, ○) and number of correctly predicted ($N_{correct}$, ●) TM helices as a function of $\Delta\Delta G_{CONH}$. As $\Delta\Delta G_{CONH}$ becomes more favorable, N_{pred} increases while $N_{correct}$ reaches a plateau. (b) Prediction accuracy Q as a function of $\Delta\Delta G_{CONH}$. Q is maximum in the vicinity of $\Delta\Delta G_{CONH} = 0$, implying that $\Delta G^{helix}_{glycyl} \approx \Delta G^{coil}_{glycyl}$. Prediction Accuracy: Single-helix prediction accuracy Q was calculated using the formula of Tusnády and Simon:[20]

$$Q = 100 \times \sqrt{\frac{N_{correct}}{N_{known}} \frac{N_{correct}}{N_{pred}}}$$

In this equation, N_{known} is the number of known TM helices in the database, N_{pred} the number predicted by the automated HΦ plot analysis, and $N_{correct}$ the number predicted correctly. Tusnády and Simon defined a successful prediction as an overlap of at least three AAs between a predicted TM segment and a known one. We examined the effect of varying the required overlap from one to nine AA. The number of prediction successes varied little over this range, typically 1-3. This behavior is expected for strong overlaps of predicted helices with known helices. But we used an overlap of three AA in order to be consistent with Tusnády and Simon.

act as an artificial TM anchor when fused to soluble alkaline phosphatase. A polyleucine segment, on the other hand, could. Their examination of the minimum hydrophobicity required for stable TM anchoring disclosed that the transition point for membrane retention occurred for 16 Ala + 5 Leu [$(LA)_3LA_{13}L$], with 14 Ala + 7 Leu [$LAL_5A_{13}L$] being completely retained. This agrees with the expectations of the WW scale: The computed free energies of transfer for these sequences are +1.75 and −1.75 kcal mol^{-1}, respectively. Peptide sequences based upon Leu-Ala repeats (net free energy = −0.75 kcal mol^{-1} per repeat) should form stable TM helices, and they do.[23,24] The situation for microsomal membranes is less clear. A systematic analysis[25] of stop-transfer activity by the introduction of hydrophobic segments of varying length and hydrophobicity into soluble interleukin 2 to form type I TM segments showed that short polyleucine segments (~10 AA) could stop translocation whereas short polyalanine segments could not. But long polyalanine segments (~20 AA) could stop translocation.

A difficulty with the interpretation of biological measurements such as these is that stop-transfer sequences perform multiple functions,[26] including both translocation halting and lateral release from the translocon into the membrane. Another difficulty is that one does not know the exact conformation of the retained peptides or their association state, whether with one another or with other membrane proteins. Finally, the kinetics of halting and release may be important.[26] Much remains to be learned about the relation between absolute TM segment hydrophobicity, stop-transfer processes, and TM helix insertion into biological membranes. The WW scale provides a rational starting point for systematic studies.

The augmented Wimley-White hydrophobicity scale

A particular advantage of the WW scale for predictions is that it includes values for the neutral forms of Asp, Glu, and His, as well as the charged forms. (But this was not utilized in the above predictions because none of the commonly used scales have this capability; see the legend to Figure 1.) In the course of developing the scale, Wimley *et al.*[18] determined the energetic consequences of salt-bridges formed between basic and acidic residues. They found that the free energy of transfer of a salt-bridge pair improved partitioning into octanol by 4.1(±0.2) kcal mol^{-1}. These additional data and the values for the neutral and charged forms of Asp, Glu, and His should permit the effects of salt-bridge formation and charge neutralization to be examined in HΦ plot analyses of specific MPs. Accounting for salt-bridges in automated analyses of large databases is presently impractical because many, if not all, salt-bridges in membrane proteins occur between helices rather than within single helices.

In Table 1, we have gathered together all of the Wimley-White data to create the augmented Wimley-White scale (aWW). Because aWW is a true thermodynamic scale obtained by direct experimental measurements of partitioning, it provides a thermodynamic foundation for understanding MP stability, including charge neutralization effects. Importantly, the $\Delta G = 0$ level, which marks the decision level for acceptance of a peak as a TM helix, is a true thermodynamic zero.

Why is ΔG^{coil}_{glycyl}, determined for *n*-octanol, a good estimate for $\Delta G^{helix}_{glycyl}$? Octanol saturated with water has a micelle-like structure[27] that allows it to readily accommodate molecules of mixed polarity, such as amino acids. We speculate that H-bonding between water, octanol hydroxyl groups, and the peptide bond are energetically equivalent to H-bonded peptide bonds of TM helices within the bilayer interior. But if wet octanol is an H-bonding solvent, are the WW values for acidic and basic residues, and salt-bridges, appropriate for proteins within the bilayer interior? We think they are. Recent thermodynamic and structural studies[28,29] indicate that basic or acidic residues buried intentionally within the hydrophobic core of soluble proteins, by site-directed mutagenesis, are accompanied by a few water molecules, which are sufficient to raise the local dielectric constant to about 12 from a ''dry'' value of about 4. This is likely true for membrane proteins as well. The interior of lactose permease is readily accessible to water and has several buried salt-bridges.[30] And a high-resolution structure of bacteriorhodopsin,[31] which has several buried charged side-chains, reveals water molecules within the protein interior. Finally, one must ask if wet octanol is appropriate for non-polar side-chains. It is. The atomic solvation parameter for non-polar side-chains in *n*-octanol[17] is −23 cal mol^{-1} Å^{-2}, virtually identical to the classical value[32] observed for pure non-polar phases.

Table 1. The augmented Wimley-White (aWW) whole-residue hydrophobicity scale,[7,17] including values for salt-bridges[18]

Amino acid residue(s)	Free energy of transfer from water (kcal mol^{-1})
Aliphatic and sulfur-containing residues	
Ala	0.50
Gly	1.15
Cys	−0.02
Ile	−1.12
Leu	−1.25
Met	−0.67
Val	−0.46
Aromatic residues	
Phe	−1.71
Trp	−2.09
Tyr	−0.71
Polar residues	
Asn	0.85
Gln	0.77
Thr	0.25
Pro	0.14
Ser	0.46
Acidic and basic residues	
Arg^+	1.81
Asp^-	3.64
Asp^0	0.43
Glu^-	3.63
Glu^0	0.11
His^+	2.33
His^0	0.11
Lys^+	2.80
Salt-bridges	
$Arg^+\cdots Asp^-$	−0.95
$Arg^+\cdots Glu^-$	−0.96
$His^+\cdots Asp^-$	−0.43
$His^+\cdots Glu^-$	−0.44
$Lys^+\cdots Asp^-$	0.04
$Lys^+\cdots Glu^-$	0.03

Practical problems associated with TM helix identification

The prediction accuracy of all the scales, including WW, is lower for the 1D_helix set compared to the 3D_helix set. The same result was obtained for augmented TM helix prediction algorithms that incorporate statistical and other data.[14] The highest prediction accuracies achieved with these methods[14] were 97 % for 3D_helix (PHDhtm[33]) and 95 % for 1D_helix (HMM[20]). The corresponding values for HΦ plots using the unaugmented WW scale are 99.1 and 97.7 % (Figure 1(b)). We suspect that the lower prediction accuracy for 1D_helix is due to an overabundance of transport proteins with buried charged groups and large extramembrane domains, which were not accounted for in our automated HΦ plot analysis. In reality, the charged groups may be neutral or have salt-bridge partners. And the extramembrane domains may have long non-polar helices that could be mistaken for TM segments.

An examination of mispredicted MPs in 3D_helix shows that in a few cases our HΦ plot algorithm fused two TM segments into one. A more clever algorithm or the incorporation of additional information, such as the inside-positive rule,[34] might solve this minor problem. The aquaporins are problematic because TM segments are formed by the end-to-end placement of short helices that are distant from one another in the amino acid sequence.[35,36] Finally, because helix D of bacteriorhodopsin is rich in Gly and contains an Asp, it was not detected in the automated HΦ plot analysis using the unaugmented WW scale. However, using the neutral form of Asp, helix D is successfully identified (Figure 3). Very specific effects such as these indicate that completely automated analysis without human input is unlikely to achieve 100 % accuracy.

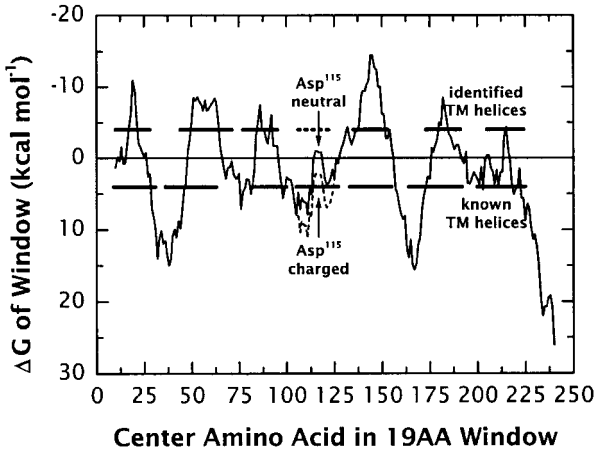

Figure 3. Hydropathy plot analysis of bacteriorhodopsin, showing the utility of being able to choose the charge state of side-chains. The horizontal grey bars indicate known TM helices,[31] and the black bars those detected by HΦ analysis. The glycine-rich fourth (D) helix is not predicted to be stable when its Asp115 is charged, as shown by the broken curved and the broken bar. When Asp115 is in protonated (continuous curve), helix D is predicted to be stable. Note, however, that even with neutral Asp115, helix D appears to be much less stable than the other predicted and known helices. This may explain why it is particularly sensitive to proteolysis.[42] The HΦ plot was performed using MPEx, located at http://blanco.biomol.uci.edu/mpex.

Identification of membrane proteins in genomes

An important issue in genomic analysis is identification of MPs in genomes, and in particular distinguishing them from soluble proteins. The problem is that simple HΦ plot analysis is likely to over-count MPs, because many soluble proteins have very non-polar subsequences. In order to explore this problem, we carried out an automated 19 AA-window HΦ analysis of 1165 soluble proteins drawn from a representative collection[37] of known 3D structure. All scales falsely identified many of these proteins as MPs. Taking the decision level as $\Delta G = 0$, the unmodified EC, KD, and GES scales, identified, respectively, 1114, 1067, and 715 of the proteins as MPs containing up to eight TM helices. Not surprisingly, the WW scale performed somewhat better by falsely predicting 499 (43%). A better comparison is to use the tuned versions of the other scales, which include optimized backbone contributions (see above). When this is done, false positives are dramatically reduced to 496 (43%), 408 (35%), and 201 (17%) for EC, KD, and GES, respectively. The good performance of the tuned GES scale is due to extremely unfavorable free energy values for charged residues. These results raise a concern about the reliability of estimates of the MP content of genomes, currently believed[38,39] to be about 30%. False positives can be reduced significantly with little effect on MP detection by using a longer window. For example, a 27 AA window reduces aWW-scale false positives to 220, but reduces the number of detected MPs by only two. This approach coupled with more sophisticated strategies[40,41,43] may allow soluble proteins to be reliably distinguished from MPs.

Epilogue

We have demonstrated the importance of the helix-backbone dehydration penalty in TM helix stability. Our estimate of the penalty appears to be accurate, but direct experimental confirmation remains important. Also important is the clarification of thermodynamic stability in the context of MP assembly by the translocon apparatus. Although further direct experimental measurements in both biological and model systems are in order, the WW hydrophobicity scale appears to provide a solid basis for predicting TM helix stability. Its augmentation with thermodynamic data for salt-bridges should extend its usefulness. Presently, however, the most important use of the aWW scale will be for topological analyses of identified MPs, especially transport proteins with buried charged groups. The aWW scale (Table 1) allows these groups to be considered explicitly in HΦ plots (Figure 3). A Web-based tool for exploring membrane protein topology using the aWW scale is available at http://blanco.biomol.uci.edu/mpex (unpublished results).

Acknowledgments

We are pleased to acknowledge Mr Michael Myers' assistance in maintaining the MPtopo database, and for his assistance in editing this manuscript. We thank Drs Alexey Ladokhin, William Wimley, Janos Lanyi, Barry Honig, Gunnar von Heijne, and Ronald Kaback for useful comments. This work was supported by the National Institutes of General Medical Sciences (GM46823).

References

1. Rose, G. D. (1978). Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature,* **272**, 586-590.
2. von Heijne, G. (1999). A day in the life of Dr K. or how I learned to stop worrying and love lysozome: a tragedy in six acts. *J. Mol. Biol.* **293**, 367-379.
3. Engelman, D. M., Steitz, T. A. & Goldman, A. (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321-353.
4. Degli Esposti, M., Crimi, M. & Venturoli, G. (1990). A critical evaluation of the hydropathy profile of membrane proteins. *Eur. J. Biochem.* **190**, 207-219.
5. White, S. H. (1994). Hydropathy plots and the prediction of membrane protein topology. In *Membrane Protein Structure: Experimental Approaches* (White, S. H., ed.), pp. 97-124, Oxford University Press, New York.

6. Lemmon, M. A. & Engelman, D. M. (1994). Specificity and promiscuity in membrane helix interactions. *Quart. Rev. Biophys.* **27**, 157-218.
7. White, S. H. & Wimley, W. C. (1999). Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319-365.
8. Popot, J.-L. & Engelman, D. M. (2000). Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* **69**, 881-922.
9. Popot, J.-L. & Engelman, D. M. (1990). Membrane protein folding and oligomerization: the 2-stage model. *Biochemistry,* **29**, 4031-4037.
10. Wimley, W. C. & White, S. H. (2000). Designing transmembrane α-helices that insert spontaneously. *Biochemistry,* **39**, 4432-4442.
11. White, S. H., Ladokhin, A. S., Jayasinghe, S. & Hristova, K. (2001). How membranes shape protein structure. *J. Biol. Chem.* **276**, 32395-32398.
12. Roseman, M. A. (1988). Hydrophobicity of the peptide C═O···H—N hydrogen-bonded group. *J. Mol. Biol.* **201**, 621-625.
13. Ben-Tal, N., Ben-Shaul, A., Nicholls, A. & Honig, B. (1996). Free-energy determinants of α-helix insertion into lipid bilayers. *Biophys. J.* **70**, 1803-1812.
14. Jayasinghe, S., Hristova, K. & White, S. H. (2001). MPtopo: a database of membrane protein topology. *Protein Sci.* **10**, 455-458.
15. Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. (1982). Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* **17**, 109-120.
16. Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
17. Wimley, W. C., Creamer, T. P. & White, S. H. (1996). Solvation energies of amino acid side-chains and backbone in a family of host-guest pentapeptides. *Biochemistry,* **35**, 5109-5124.
18. Wimley, W. C., Gawrisch, K., Creamer, T. P. & White, S. H. (1996). A direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proc. Natl Acad. Sci. USA,* **93**, 2985-2990.
19. Hristova, K. & White, S. H. (1998). Determination of the hydrocarbon core structure of fluid dioleoylphosphocholine (DOPC) bilayers by X-ray diffraction using specific bromination of the double-bonds: effect of hydration. *Biophys. J.* **74**, 2419-2433.
20. Tusnády, G. E. & Simon, I. (1998). Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **283**, 489-506.
21. Wallace, B. A. (1998). Recent advances in the high resolution structures of bacterial channels: Gramicidin A. *J. Struct. Biol.* **121**, 123-141.
22. Chen, H. & Kendall, D. A. (1995). Artificial transmembrane segments: requirements for stop transfer and polypeptide orientation. *J. Biol. Chem.* **270**, 14115-14122.
23. Zhang, Y.-P., Lewis, R. N. A. H., Henry, G. D., Sykes, B. D., Hodges, R. S. & McElhaney, R. N. (1995). Peptide models of helical hydrophobic transmembrane segments of membrane proteins. 1. Studies of the conformation, intrabilayer orientation, and amide hydrogen exchangeability of Ac-K_2-$(LA)_{12}$-K_2-amide. *Biochemistry,* **34**, 2348-2361.
24. Zhang, Y.-P., Lewis, R. N. A. H., Hodges, R. S. & McElhaney, R. N. (1995). Peptide models of helical hydrophobic transmembrane segments of membrane proteins. 2. Differential scanning calorimetric and FTIR spectroscopic studies of the interaction of Ac-K_2-$(LA)_{12}$-K_2-amide with phosphatidylcholine bilayers. *Biochemistry,* **34**, 2362-2371.
25. Kuroiwa, T., Sakaguchi, M., Mihara, K. & Omura, T. (1991). Systematic analysis of stop-transfer sequence for microsomal membrane. *J. Biol. Chem.* **266**, 9251-9255.
26. Duong, F. & Wickner, W. (1998). Sec-dependent membrane protein biogenesis: SecYEG, preprotein hydrophobicity and translocation kinetics control the stop-transfer function. *EMBO J.* **17**, 696-705.
27. Franks, N. P., Abraham, M. H. & Lieb, W. R. (1993). Molecular organization of liquid *n*-octanol: an X-ray diffraction analysis. *J. Pharm. Sci.* **82**, 466-470.
28. García-Moreno, E. B., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S. & Stites, W. E. (1997). Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophys. Chem.* **64**, 211-224.
29. Dwyer, J. J., Gittis, A. G., Karp, D. A., Lattman, E. E., Spencer, D. S., Stites, W. E. & García-Moreno, E. B. (2000). High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **79**, 1610-1620.
30. Kaback, H. R. & Wu, J. H. (1999). What to do while awaiting crystals of a membrane transport protein and thereafter. *Acc. Chem. Res.* **32**, 805-813.
31. Luecke, H., Schobert, B., Richter, H.-T., Cartailler, J.-P. & Lanyi, J. K. (1999). Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899-911.
32. Reynolds, J. A., Gilbert, D. B. & Tanford, C. (1974). Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Natl Acad. Sci. USA,* **71**, 2925-2927.
33. Rost, B., Casadio, R., Fariselli, P. & Sander, C. (1995). Transmembrane helices predicted at 95% accuracy. *Protein Sci.* **4**, 521-533.
34. von Heijne, G. (1992). Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487-494.
35. Mitsuoka, K., Murata, K., Walz, T., Hirai, T., Agre, P., Heymann, J. B. *et al.* (1999). The structure of aquaporin-1 at 4.5-Å resolution reveals short α-helices in the center of the monomer. *J. Struct. Biol.* **128**, 34-43.
36. Fu, D., Libson, A., Miercke, L. J. W., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000). Structure of a glycerol-conducting channel and the basis for its selectivity. *Science,* **290**, 481-486.
37. Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522-524.
38. Arkin, I. T., Brünger, A. T. & Engelman, D. M. (1997). Are there dominant membrane protein families with a given number of helices? *Proteins: Struct. Funct. Genet.* **28**, 465-466.
39. Wallin, E. & von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029-1038.
40. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-5801.
41. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. (1984). Analysis of membrane and surface protein

sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179**, 125-142.
42. Fimmel, S., Choli, T., Dencher, N. A., Büldt, G. & Wittmann-Liebold, B. (1989). Topography of surface-exposed amino acids in the membrane protein bacteriorhodopsin determined by proteolysis and micro-sequencing. *Biochim. Biophys. Acta,* **978**, 231-240.
43. White, S. H. & Jacobs, R. E. (1990). Observations concerning topology and locations of helic ends of membrane proteins of known structure. *J. Membrane Biol.* **115**, 145-158.

Edited by G. von Heijne

(Received 12 June 2001; received in revised form 10 August 2001; accepted 13 August 2001)