The Identification of Transmembrane Helices in the Sequences of Membrane Proteins Using a Computationally-Derived Hydrophobicity Scale

Lisa Bourla[1] **Tidhar Seifer**[1]

Barry Honig[2] **Nir Ben-Tal**[1]
bental@ashtoret.tau.ac.il

[1] Department of Biochemistry, Tel Aviv University, Ramat Aviv 69978, Israel
[2] Department of Biochemistry and Molecular Biophysics Columbia University, 630 West 168th St., NY, NY 10032, USA

Keywords: transmembrane helix prediction, hydropathy scale, continuum solvent models, dynamic programming algorithm

1 Introduction

All evidence suggests that about 20-30% of all proteins are integral membrane proteins. The overexpression and crystallization of membrane proteins is, however, difficult and thus, three-dimensional (3D) structures have been determined for only several of the over 13,000 sequences of membrane proteins currently available in Swiss-Prot. Most of the membrane proteins of known structure are α-helix bundles that span the membrane. All experimental observations thus far suggest that transmembrane helices are autonomous folding domains (reviewed in [1]). Thus, these observations solidly support the development of algorithmic tools for predicting the 3D structures of membrane proteins in three steps: (i) Prediction of the locations of the transmembrane helices in the sequence. (ii) Determination of the relative orientation of each transmembrane helix with respect to the others in the lipid bilayer. (iii) Building loops connecting the helices. The present study describes the development of a computationally derived hydrophobicity scale for the transfer of amino acids from water to bilayers in the context of an α-helix, and its implementation in a threading method to identify the locations of transmembrane spans in the sequences of membrane proteins.

2 Methodology and Results

We have used continuum solvent models to calculate the transfer free energies of polyalanine α-helices from the aqueous phase into lipid bilayers and the results were in very good agreement with experimental data [2]. In this study we used the same methodology to derive a hydrophobicity scale of the transfer free energies of the twenty amino acids from the aqueous phase into the lipid bilayer in the context of a polyalanine α-helix. We then used the scale in a dynamic programming algorithm to identify transmembrane spans in the sequences of membrane proteins [3]. The algorithm is based on a summation of the free energies of transfer of the amino acids over a sliding window.

We tested the methodology on a set of over 140 bacterial and eukaryotic integral membrane protein sequences. The set, which is based on the set of Tusnady and Simon [4] with a few additions, including proteins that have a reliable experimentally determined topology. The set is used as a standard benchmark to test algorithms for topology prediction of membrane proteins. Using our scale, we correctly identified the tramsmembrane spans in about 60% of the proteins in the set. Comparison of our results with results from predictions which used the same algorithm but with previously existing scales shows that our scale gives the best performance, presumably because it was derived for amino acids in the context of an α-helix. The model is fairly successful in predicting transmembrane helices in proteins containing single and double span proteins. Its performance for this type of protein is better than that of statistical methods. However, its performance is somewhat worse than that of the statistics-based methods for proteins with three or more transmembrane spans, so that its overall performance is not as good. This behavior is to be expected because, although protein-membrane interactions are taken into account in the model, the pair-wise interactions between residues in adjacent helices are not.

3 Discussion

The availability of membrane proteins with known high-resolution structures enables a closer examination of our model. The model successfully predicted the transmembrane helices of all these proteins with one exception, helix number 3 in bacteriorhodopsin. This helix is the least exposed to the lipid. An analysis of the amino acid distribution along this helix revealed that it is amphipathic and that it is oriented so that its hydrophilic side faces the interior of the protein and its hydrophobic face is exposed to the lipid.

Based on the structural analysis, we improved the prediction capabilities of the algorithm by introducing a possibility to search for transmembrane helices that are mildly amphipathic. This is, of course, relevant in the case of proteins with multiple transmembrane spans. The idea is that a (narrow) hydrophilic face of a transmembrane helix may be buried inside the helix bundle, while the rest of the helix is lipid-exposed. The residues in the hydrophilic face should obviously be omitted from the sum over the sliding window. Our algorithmic implementation of the idea is based on the approximation that transmembrane spans are canonical α-helices that can be represented on a helical wheel, i.e. sequential resides should be 100° away from each other on the wheel. The algorithm involves the screening of 36 possibilities of amphipathic helices for each sliding window.

This modified algorithm detects the location of helix number 3 in bacteriorhodopsin. However, the search for amphipathic helices increases the number of false positive predictions, as well. We are currently conducting tests for the optimal value of the angular hydrophilic slice. Setting the width of the hydrophilic slice to 25° resulted in us correctly identifing all the transmembrane spans (without false positives) in 80% of the proteins in the set.

References

[1] Popot, J.L., Integral membrane protein structure: transmembrane α-helices as autonomous folding domains, *Curr. Op. Struct. Biol.*, 3:532–540, 1993.

[2] Ben-Tal, N., Ben-Shaul, A, Nicholls, A., and Honig, B., Free-energy determinants of α-helix insertion into lipid bilayers, *Biophys. J.*, 70:1803–1812, 1996.

[3] Jones, D.T., Taylor, W.R., and Thornton, J.M., A model recognition approach to the prediction of all-helical membrane protein structure and topology, *Biochemistry*, 33:3038–3049, 1994.

[4] Tusnady, G.E. and Simon, I., Principles governing amino acid composition of integral membrane proteins, *J. Mol. Biol.*, 283(2):489–506, 1998.