Do protein–lipid interactions determine the recognition of transmembrane helices at the ER translocon?

S.H. White* and G. von Heijne†
*Department of Physiology and Biophysics and Program in Macromolecular Structure, University of California at Irvine, Irvine, CA 92697, U.S.A., and †Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden

Abstract

Membrane-protein integration, folding and assembly processes in vivo depend on complex targeting, translocation, chaperoning, and sorting machineries that somehow read the ‘molecular code’ built into the nascent polypeptide, ultimately producing a properly folded protein integrated into the correct target membrane. Although the main molecular constituents and the basic mechanistic principles of many of these machines are known in outline, the codes remain poorly defined and there is little quantitative information on how protein sequence affects the final structure of membrane proteins. By carefully designing model protein constructs, we have derived the first true biological hydrophobicity scale and have been able to get a first impression of how the position of a given type of residue within a transmembrane segment affects its ability to promote membrane insertion.

Introduction

The exposed surface of integral membrane proteins generally seems to reflect quite accurately the physicochemical properties of the surrounding lipid bilayer. As a general rule [1,2], there is a central section (approx. 20 Å wide; 1 Å = 0.1 nm) rich in aliphatic residues that interfaces with the hydrocarbon lipid tails, sandwiched between two sections (each approx. 10 Å wide) with a high content of tryptophan and tyrosine residues, which are known to interact preferentially with lipid headgroups. Finally, surface-exposed polar and charged residues appear approx. 20 Å or more away from the centre of the membrane, where they can easily reach into the aqueous environment.

In terms of secondary structure, the central 20 Å is composed almost completely of transmembrane α-helices; the region around ±25 Å contains both transmembrane and surface-localized α-helices as well as irregular structure, but little or no β-structure; and finally, at ≥35 Å from the membrane centre, the secondary-structure composition approximates that found in globular water-soluble proteins [3].

From the perspective of membrane-protein structure, the picture seems clear: membrane proteins are optimized to interact favourably with lipid bilayers. But there is a paradox here: membrane proteins do not as a rule insert spontaneously into cellular membranes, but instead are processed co-translationally through a complex membrane-embedded molecular machine, a translocon, which ensures the recognition and proper membrane insertion of the transmembrane parts [4]. So do translocons in fact recognize precisely the same molecular characteristics in the nascent membrane protein that are responsible for the ultimate matching between the protein and the membrane? And if so, how is this achieved?

Recognition of transmembrane helices by the ER (endoplasmic reticulum) translocon

Our initial approach to these questions has been to challenge the archetypal translocon, the Sec61 complex in the ER membrane, with large panels of engineered polypeptide segments designed to explore the importance for membrane insertion of different kinds of residues located in different positions relative to the lipid bilayer. Taking advantage of a robust and easy-to-use in vitro transcription–translation system, supplanted with ER-derived microsomal membrane vesicles and a well-characterized model membrane protein (Figure 1), we have been able to derive a first quantitative ‘biological’ hydrophobicity scale, which reflects the contribution to the total free energy of membrane insertion of the 20 natural amino acids when placed in the middle of a transmembrane helix [5] (Figure 2).

The biological scale correlates well with the statistical properties of membrane-protein structures: aliphatic amino acids favour membrane insertion (ΔG^\text{\scriptscriptstyle{app}} < 0, where aa is amino acid), polar amino acids (including the polar aromatic residues tryptophan and tyrosine) are somewhat unfavourable, and charged amino acids are very unfavourable. As an example, it takes four to five leucine residues to balance the unfavourable contribution from a single centrally placed lysine. The structure of the biological scale, and the observation that the scale is to a good approximation additive ([6], but see [7]), already suggests that the recognition of transmembrane helices, even within the context of
The model protein *Escherichia coli* Lep has two N-terminal transmembrane segments (TM1 and TM2) and a large luminal domain (P2). H-segments (grey) are inserted into the P2 domain, where they are flanked by GGPG tetrapeptides and by two glycosylation acceptor sites (G1 and G2). If H integrates across the membrane, only G1 is glycosylated (left), otherwise both G1 and G2 are glycosylated (right). The apparent free energy of insertion of an H-segment is defined as $\Delta G_{\text{app}} = RT \ln(f_1g/f_2g)$, where $f_1g$ and $f_2g$ are the fractions of singly and doubly glycosylated molecules respectively. Reproduced with permission from [5].

For each residue, its contribution to $\Delta G_{\text{app}}$ when placed in the middle of a 19-residues-long H-segment with the flanking sequences GGPG...GPGG is shown. Reproduced with permission from [5].

Translocon-mediated insertion, may be based on protein–lipid rather than protein–protein interactions.

This idea receives even stronger support from an analysis of the positional dependence of the $\Delta G_{\text{app}}$ values for different residues. Two examples are shown in Figure 3(A). Mildly polar residues, such as glycine, display little positional variation, whereas the contribution from charged residues, such as arginine, depends dramatically on their position within the transmembrane helix.

A particularly interesting case is proline, the classic helix-breaker, which also shows a very strong positional dependence (Figure 3B). Strikingly, the effect of proline is clearly asymmetric: proline is not particularly unfavourable in the three most N-terminal positions, but it has a much stronger effect when placed in the three most C-terminal positions.
positions (as shown by the arrow). In soluble proteins, it is well known that proline can be accommodated in the most N-terminal, but not in the corresponding C-terminal, positions of an α-helix. The positional dependence of \( \Delta G^{\text{Pro}}_{\text{app}} \) is thus most easily explained if one assumes that the transfer of a transmembrane segment from the translocon into the surrounding lipid bilayer requires that the segment folds into a helical structure.

Based on ‘symmetric scans’, we have also quantified the positional dependence of leucine as well as the aromatic residues phenylalanine, tryptophan and tyrosine (Figure 3C). The most telling result of these experiments is the difference between leucine and phenylalanine on the one hand, and tryptophan and tyrosine on the other. Neither leucine nor tryptophan show a strong positional dependence in their \( \Delta G^{\text{aa}}_{\text{app}} \) values (the undulatory aspect of the curves seems to be related to the amphiphilicity of the segments [5]); in contrast, tryptophan and tyrosine are quite unfavourable when placed near the middle of the transmembrane helix, but become quite favourable when moved towards the ends of the helix, reflecting their distribution in the known membrane-protein structures.

Finally, as a critical test of the notion that the \( \Delta G^{\text{aa}}_{\text{app}} \) scale is additive, we have analysed the so-called voltage-sensor helix (S4) from the KvAP voltage-regulated K\(^+\) channel ([6], but see [7]). This remarkable helix contains no less than four arginines; however, experimentally we find that S4 is on the brink of being able to form a stable transmembrane helix (\( \Delta G_{\text{app}} \) is approx. 0.5 kcal/mol (1 cal = 4.184 J)). Simply summing the individual residue contributions to \( \Delta G_{\text{app}} \) using the ‘biological’ hydrophobicity scale predicts that S4 should not insert at all (\( \Delta G_{\text{app}} \) is approx. 3.9 kcal/mol). However, if the positional variations shown for arginine and glycine in Figure 3 are taken into account, the calculated value for \( \Delta G_{\text{app}} \) is approx. 0.9 kcal/mol, not too far from the experimental value.

Taken together, the experiments carried out thus far provide a picture of what it takes to design a segment that will be recognized as a transmembrane helix by the ER translocon that is remarkably consistent with the statistical distributions seen in the high-resolution membrane-protein structures. We therefore suggest that the Sec61 translocon in the ER dynamically opens up laterally towards the surrounding lipid bilayer, thereby providing nascent protein chains that are in transit direct access to the lipid environment, and that this direct protein–lipid interaction is the basis for the recognition process. Although this proposal is not inconsistent with the three-dimensional structure of the Sec61 translocon [8], clearly more work is needed both to substantiate it further and to arrive at a more detailed molecular-level description of the process.

Molecular dynamics simulations of transmembrane helices

With the quantitative information on transmembrane-helix integration offered by the experiments described above, a natural next step is to see whether a more detailed understanding might be arrived at through molecular dynamics simulations of bilayer-embedded transmembrane helices. As a first step in this direction, we recently analysed the S4 helix from the KvAP K\(^+\) channel in a POPC (1-palmitoyl-2-oleoyl phosphatidylcholine) bilayer (J.A. Freites, D.J. Tobias, G. von Heijne and S.H. White, unpublished work). The simulation revealed a stabilizing hydrogen-bonded network of water and lipid phosphates around the arginines that reduces the effective thickness of the bilayer hydrocarbon core to approx. 10 Å in the vicinity of the helix. This result provides a graphic illustration of how continuum electrostatic models, in which the bilayer is modelled as a homogeneous hydrocarbon slab with bulk properties, break down on the atomic scale. Instead, the lipid dynamics apparently permits lipids in the immediate vicinity of a transmembrane helix to adopt conformations that place their headgroups in contact with buried peptide charges, thereby providing a scaffold for water penetration. While it is still difficult to estimate the quantitative energetics of such distorted bilayer states, it is not unlikely that their appearance may explain the strong positional dependence of the \( \Delta G^{\text{aa}}_{\text{app}} \) values seen in the biological experiments.

Conclusions

The development of an in vitro system that closely mirrors the in vivo mechanism of membrane-protein insertion into the ER membrane has made it possible to put the study of membrane-protein assembly on a quantitative footing. The first studies using this system strongly suggest that the recognition of transmembrane helices by the ER translocon critically involve direct protein–lipid interactions, although we assume that potentially revealing second-order effects, which depend on details of both the translocon itself and the particular membrane protein being synthesized, will be identified in the future.

The observation that the \( \Delta G^{\text{aa}}_{\text{app}} \) values for charged residues, in particular, vary strongly with position in the transmembrane helix is an indication that detailed atomic-level modelling of protein–lipid interactions will be necessary to arrive at a realistic picture of protein–lipid interaction energies; continuum models that ignore the fine structure of the lipid bilayer do not suffice. This is quite clear already from fairly standard molecular dynamics simulations of transmembrane helical peptides, but much remains to be done in this area before reliable energetics can be calculated from first principles.

This work was supported by grants from the National Institute of General Medical Sciences and the National Center for Research Resources to S.H.W., and the Swedish Cancer Foundation, the Swedish Research Council, and the Marianne and Marcus Wallenberg Foundation to G.v.H.

References

7 Erratum (2005) Science 308, 1743

Received 21 July 2005