Recognition of transmembrane helices by the endoplasmic reticulum translocon

Tara Hessa[1], Hyun Kim[1], Karl Bihlmaier[1]*, Carolina Lundin[1], Jorrit Boekel[1], Helena Andersson[3], IngMarie Nilsson[1], Stephen H. White[2] & Gunnar von Heijne[1]

[1]*Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden*
[2]*Department of Physiology and Biophysics and the Program in Macromolecular Structure, University of California at Irvine, Irvine, California 92697-4560, USA*
[3]*Karolinska Institutet, Department of Bioscience at NOVUM, SE-141 57 Huddinge, Sweden*

* Present address: Institute of Biochemistry, ETH Zurich, CH-8093 Zurich, Switzerland

Membrane proteins depend on complex translocation machineries for insertion into target membranes. Although it has long been known that an abundance of nonpolar residues in transmembrane helices is the principal criterion for membrane insertion, the specific sequence-coding for transmembrane helices has not been identified. By challenging the endoplasmic reticulum Sec61 translocon with an extensive set of designed polypeptide segments, we have determined the basic features of this code, including a 'biological' hydrophobicity scale. We find that membrane insertion depends strongly on the position of polar residues within transmembrane segments, adding a new dimension to the problem of predicting transmembrane helices from amino acid sequences. Our results indicate that direct protein–lipid interactions are critical during translocon-mediated membrane insertion.

Integral membrane proteins account for 20–30% of all genes in both prokaryotic and eukaryotic organisms[1] and are involved in a host of biological functions such as signal transduction, solute and macromolecular transport across membranes, cell–cell interactions and nerve conduction. The great majority of integral membrane proteins belong to the helix-bundle class[2], that is, their basic architecture is one of tightly packed transmembrane (TM) α-helices.

In eukaryotic cells, most helix-bundle membrane proteins insert co-translationally and fold in the endoplasmic reticulum (ER) membrane[3]. Insertion is mediated by the Sec61 translocon, a hetero-oligomeric protein-conducting channel[4]. Individual TM helices seem to follow an ordered insertion pathway, in which they pass from the tunnel in the large ribosomal subunit into the Sec61 translocon channel and then exit the channel laterally into the surrounding lipid[5,6]. The recently determined X-ray structure of an archaeal SecY translocon (homologous to the eukaryotic Sec61 translocon) suggests that segments in a translocating polypeptide can move laterally into the ER membrane through a gate located between helices 2b–3 and 7–8 in the Sec61 α-subunit to form TM helices[7].

A fundamental question concerns the design of the 'code' that determines whether or not a polypeptide segment is recognized by the translocon for integration into the ER membrane[8]. Here, we approach this question by comparing extensive data on TM helix integration obtained from quantitative studies of the translocon-mediated pathway with data obtained from biophysical studies of well defined model systems[9,10] and from statistical analyses of membrane protein structures[11,12]. Overall, our results strongly indicate that direct interactions between the nascent polypeptide in transit through the ER membrane and the lipid bilayer surrounding the Sec61 translocon are critical in the recognition of TM helices.

Connecting biological and biophysical principles

The high content of nonpolar residues in TM segments is a clear indication that hydrophobicity is an important component of the TM-sequence code read by the translocon. To assess the importance of hydrophobicity, we challenged the Sec61 translocon with a large set of systematically designed TM sequences and measured the efficiency of membrane integration for each one. To the extent that quantitative data generated in this way can be decomposed into contributions from individual residues, the output from the biological system can be directly compared with biophysical measurements.

Precise quantification of the biological process is essential for such comparisons. We use an *in vitro* assay[13] for quantifying the efficiency of membrane integration of designed TM segments into dog pancreas rough microsomes (RMs). Briefly, a segment to be tested (H-segment) is engineered into the luminal P2 domain of the integral membrane protein leader peptidase (Lep), where it is flanked by two acceptor sites for N-linked glycosylation (Fig. 1a). The degree of membrane integration of the H-segment is quantified from SDS–PAGE gels (Fig. 1b) by measuring the fraction of singly (f_{1g}) versus singly plus doubly (f_{2g}) glycosylated Lep molecules, $p = f_{1g}/(f_{1g} + f_{2g})$. The data can also be expressed as an apparent equilibrium constant, $K_{app} = f_{1g}/f_{2g}$, between the membrane integrated and non-integrated forms. Although this expression is not meant to imply that the membrane integration process reflects a true thermodynamic equilibrium (but this does seem to be a good first approximation, see below), it has the advantage that the results can be converted to apparent free energies, $\Delta G_{app} = -RT\ln K_{app}$, for direct comparison with biophysical data.

We have studied H-segments with the general design GGPG-X_{19}-GPGG, in which the flanking tetrapeptides are included to 'insulate' the central 19-residue stretch from the surrounding sequence. The flanking tetrapeptides were chosen to be indifferent with respect to charge and other specific interactions that might occur in the membrane and/or translocon, but to have a low probability of secondary structure formation.

A biological hydrophobicity scale

The Wimley–White water/octanol free energy scale[10,14] predicts that H-segments composed of n leucine and $(19 - n)$ alanine residues will be stably inserted across the membrane for $n \approx 5$–6[15]. We thus

began by testing H-segments of the design GGPG-(L_nA_{19-n})-GPGG with $n = 0$–7. Results obtained when 2–4 Leu residues are incorporated in various positions along the sequence are shown in Fig. 1c. There is an overall linear relationship between n and ΔG_{app}, a simple outcome consistent with energy-additivity. Furthermore, the probability of insertion, p, conforms to a Boltzmann distribution (Fig. 1d). The Boltzmann distribution shows that translocon-mediated insertion has the appearance of an equilibrium process, supporting the equilibrium approximation behind the calculation of ΔG_{app}. Additionally, Fig. 1c shows that for a given n, there is a variation in ΔG_{app} of ± 0.2–0.3 kcal mol^{-1} around the mean when the positions of the Leu residues are changed. Thus, total hydrophobicity is not the sole determinant of membrane insertion; residue position also has an important influence.

Ignoring for the moment the positional variation, the data in Fig. 1c can be fitted by the expression:

$$\Delta G_{app} = -0.66n + 2.14 = n\Delta G^{Leu}_{app} + (19 - n)\Delta G^{Ala}_{app} + \Delta G^{flank}_{app}$$
$$= n(\Delta G^{Leu}_{app} - \Delta G^{Ala}_{app}) + 19\Delta G^{Ala}_{app} + \Delta G^{flank}_{app} \quad (1)$$

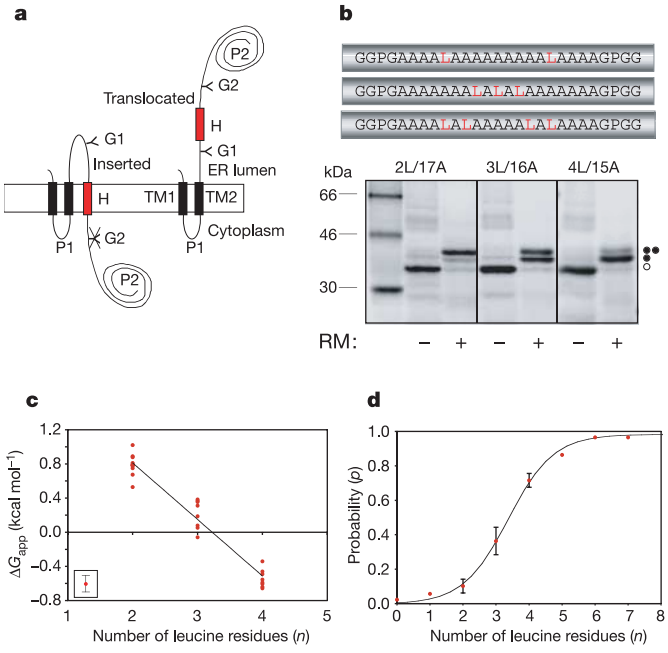

Figure 1 Integration of H-segments into the microsomal membrane. **a**, Wild-type Lep has two N-terminal TM segments (TM1 and TM2) and a large luminal domain (P2). H-segments were inserted between residues 226 and 253 in the P2-domain. Glycosylation acceptor sites (G1 and G2) were placed in positions 96–98 and 258–260, flanking the H-segment. For H-segments that integrate into the membrane, only the G1 site is glycosylated (left), whereas both the G1 and G2 sites are glycosylated for H-segments that do not integrate in the membrane (right). **b**, Membrane integration of H-segments with the Leu/Ala composition 2L/17A, 3L/16A and 4L/15A. Plasmids encoding the Lep/H-segment constructs were transcribed and translated *in vitro* in the presence (+) and absence (−) of dog pancreas rough microsomes (RM). Bands of unglycosylated protein are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. **c**, ΔG_{app} values for H-segments with 2–4 Leu residues. The average standard deviation in the individual ΔG_{app} determinations is shown in the boxed insert. Individual points for a given n show ΔG_{app} values obtained when the position of Leu is changed. **d**, Mean probability of insertion (p) for H-segments with $n = 0$–7 Leu residues. The curve shown represents the best-fit Boltzmann distribution. For $n = 2$–4, mean values and standard deviations were calculated by averaging the data in **c**. For $n = 0, 1, 5$–7, only single H-segments with the following compositions were used (flanked by GGPG…GPGG in all cases): $(A)_{19}$, $(A)_9L(A)_9$, $(A)_4LALAALAALAL(A)_4$, $(A)_4(LA)_5L(A)_4$, ALAALALAALAALALAALA.

where the last term accounts for the contribution to ΔG_{app} from the two tetrapeptides flanking the central 19-residue segment. The average $\Delta\Delta G^{Ala\rightarrow Leu}_{app}$ for an Ala → Leu replacement is thus -0.7 kcal mol^{-1}. On the assumption that $\Delta G^{flank}_{app} = 0$ (see below), we can further calculate individual ΔG^{aa}_{app} values: $\Delta G^{Ala}_{app} = 0.1$ kcal mol^{-1} and $\Delta G^{Leu}_{app} = -0.6$ kcal mol^{-1}.

We next determined ΔG^{aa}_{app} values for the remaining 18 naturally occurring amino acids when placed in the middle of the 19-residue stretch. The quantification is maximally sensitive for H-segments

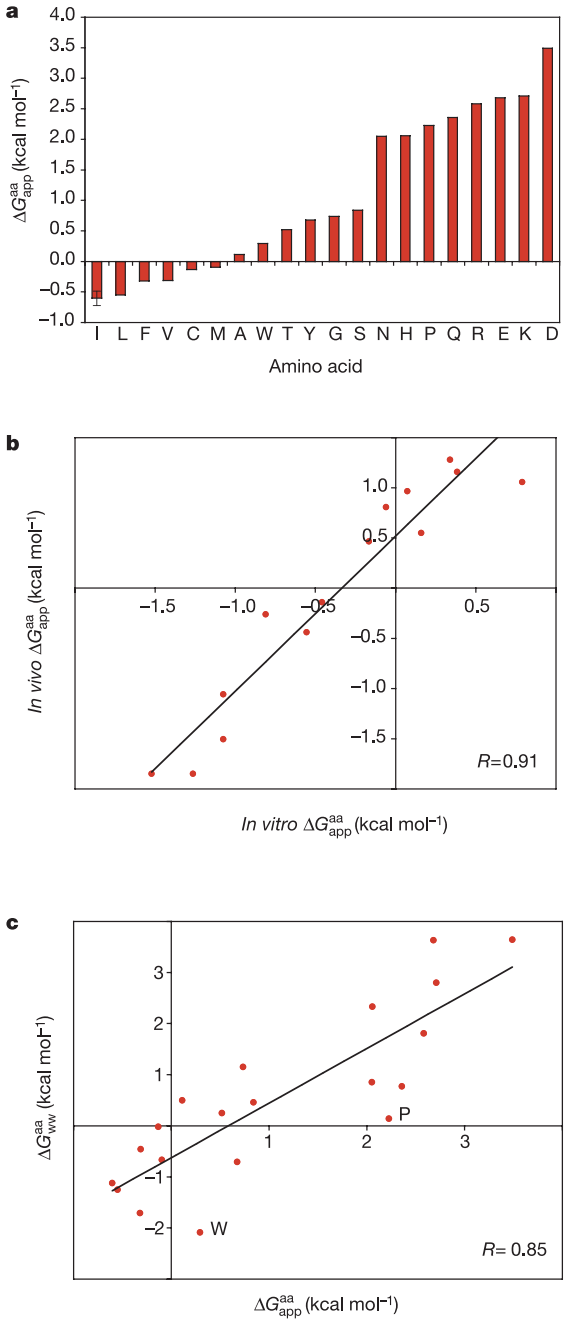

Figure 2 Biological and biophysical ΔG^{aa} scales. **a**, ΔG^{aa}_{app} scale derived from H-segments with the indicated amino acid placed in the middle of the 19-residue hydrophobic stretch (see Supplementary Information S1 for details). The bar indicates the standard deviation in the determination of ΔG^{Ile}_{app}; the standard deviation for all other amino acids is similar. **b**, Correlation between ΔG^{aa}_{app} values measured *in vivo* and *in vitro*. See Supplementary Information S3 for details on the H-segments used. **c**, Correlation between the ΔG^{aa}_{app} scale and the Wimley–White water/octanol free energy scale[14] (ΔG^{aa}_{ww}).

with ΔG_{app} values close to zero ($p \approx 0.5$ in Fig. 1d); therefore, for each kind of residue we balanced the contribution from the central residue by varying the number of Leu residues until an H-segment with ΔG_{app} in the range -1.2 to 1.2 kcal mol^{-1} was found. Based on these data, a full ΔG^{aa}_{app} scale was calculated using a stepwise procedure (described in Supplementary Information S1). As seen in Fig. 2a, Ile, Leu, Phe and Val have $\Delta G^{aa}_{app} < 0$ and thus promote membrane insertion, Cys, Met and Ala have $\Delta G^{aa}_{app} \approx 0$, and all polar and charged residues have $\Delta G^{aa}_{app} > 0$. A charged residue can be tolerated in the centre of a TM segment if its unfavourable cost is offset by a sufficiently large number of nonpolar residues, as expected for thermodynamic equilibrium. Related H-segments in which either one or two Ala residues have been changed to Gly, Ser, Trp or Tyr all have $\Delta\Delta G^{Ala\rightarrow X}_{app} \approx 0.5\Delta\Delta G^{2Ala\rightarrow 2X}_{app}$ (see Supplementary Information S2), suggesting that the ΔG^{aa}_{app} scale is approximately additive. To ensure that the *in vitro* results are relevant to the *in vivo* situation, selected constructs were also expressed *in vivo* in BHK cells. On average, the ΔG_{app} values changed by only 0.5 kcal mol^{-1}, corresponding to an increase in the individual ΔG^{aa}_{app} values by 0.03 kcal mol^{-1} compared to the *in vitro* results (Fig. 2b).

The 'biological' ΔG^{aa}_{app} scale clearly has much in common with hydrophobicity scales derived from biophysical measurements[16,17] and with statistical analyses of the lipid-exposed parts of high-resolution membrane protein structures[11,12]. The correlation between the ΔG^{aa}_{app} scale and the Wimley–White water/octanol free energy scale[14] is shown in Fig. 2c. Considering the complexity of the biological system, the two scales correlate surprisingly well: the linear fit has a slope of 1.1 and the origins of the two scales coincide within 0.5 kcal mol^{-1}. The only clear amino acid outliers are Trp (W) and Pro (P), which are both more hydrophobic on the Wimley–White scale. The correlation between the two scales (and between the ΔG^{aa}_{app} scale and other biophysical scales, data not shown) immediately suggests that the recognition of TM segments by the translocon involves direct interaction between the TM segment and the surrounding lipid[18].

Can the contribution from the flanking tetrapeptides (ΔG^{flank}_{app}) significantly affect these conclusions? To address this question, we first lengthened the flanking stretches of Gly residues stepwise from GGPG-X_{19}-GPGG to GGGGGGPG-X_{19}-GPGGGGGG for an H-segment with 3 Leu and 16 Ala residues (3L/16A). For this series, the ΔG_{app} values vary by no more than ± 0.2 kcal mol^{-1} (see Supplementary Information S4), demonstrating that residues in the Lep P2 domain outside the H-segment have little influence on the results. We also replaced the GGPG-X_{19}-GPGG flanks by NNPN-X_{19}-NPNN and tested H-segments containing three or

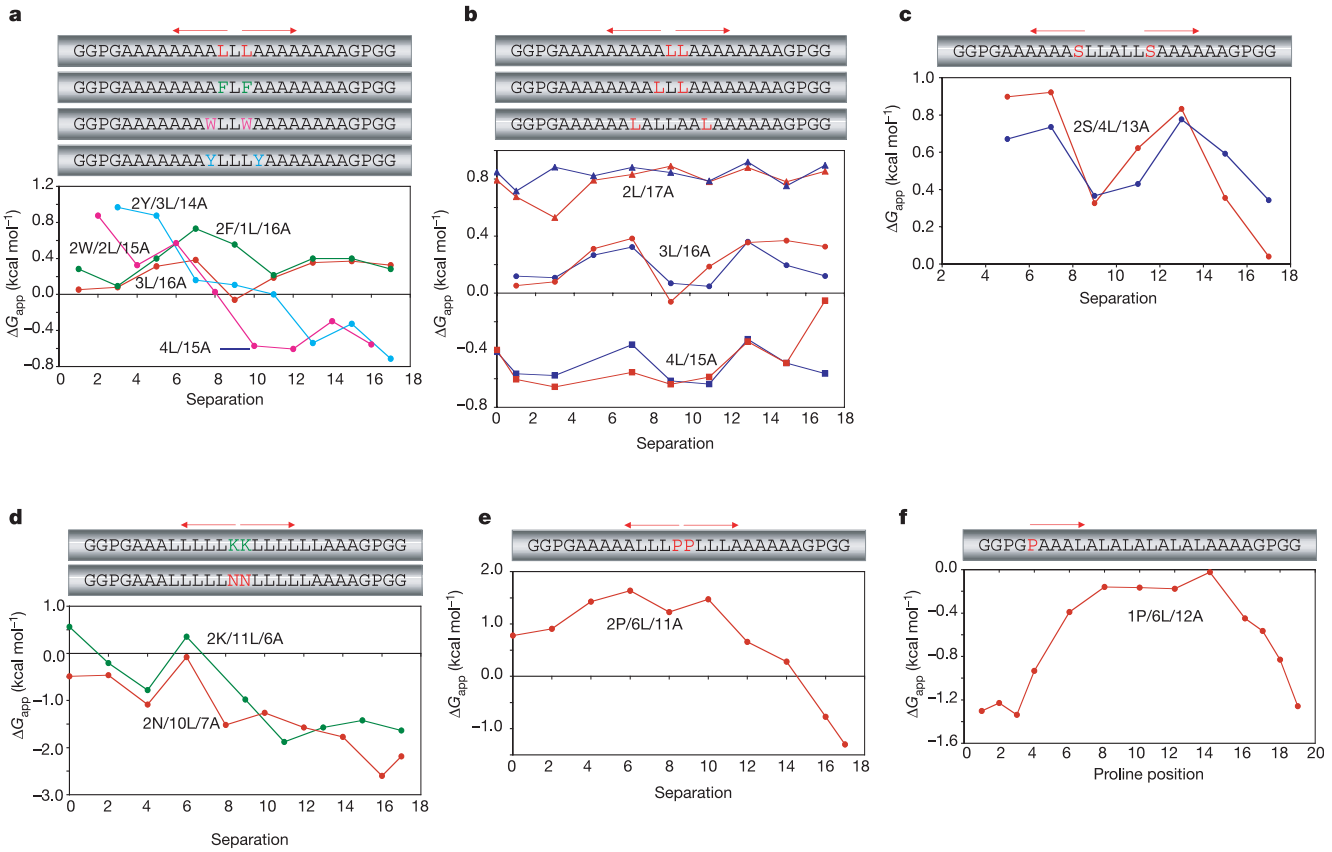

Figure 3 Positional dependencies in ΔG_{app}. **a**, Symmetrical H-segment scans with pairs of Leu (red), Phe (green), Trp (pink) or Tyr (light blue) residues. The Leu scan is based on symmetrical 3L/16A H-segments with a Leu-Leu separation of one residue (sequence shown at the top; the two red Leu residues are moved symmetrically outwards) up to a separation of 17 residues. For the Phe scan, the composition of the central 19-residues of the H-segments is 2F/1L/16A, for the Trp scan it is 2W/2L/15A, and for the Tyr scan it is 2Y/3L/14A. The ΔG_{app} value for the 4L/15A H-segment GGPGAAALAALAAAAALAALAAAGPGG is also shown (dark blue). **b**, Red lines show ΔG_{app} values for symmetrical scans of 2L/17A (triangles), 3L/16A (circles), and 4L/15A (squares) H-segments. Blue lines show the ΔG_{app} values predicted from least-square linear fits between the observed data and the hydrophobic moment ($\delta = 100°$) of the central 19-residue segments (see Methods). The predicted values all have the form $\Delta G_{app} = 0.17\mu + h_i$, where μ is the hydrophobic moment and h_i is the extrapolated ΔG_{app} value for an H-segment of the given composition with $\mu = 0$. **c**, Same as **b** but for a symmetrical scan with pairs of Ser residues in H-segments with the composition 2S/4L/13A. **d**, Symmetrical scans with pairs of Lys (green) and Asn (red) residues in H-segments with the overall compositions 2K/11L/6A and 2N/10L/7A. **e**, Symmetrical scan with a pair of Pro residues in H-segments with the composition 2P/6L/11A. **f**, Scan with a single Pro residue in H-segments with the composition 1P/6L/12A. The position of the Pro residue in the 19-residue central hydrophobic stretch is shown on the *x*-axis.

four Leu residues. On average, ΔG_{app} increases by only 0.5 kcal mol^{-1} when the six Gly residues in the flanking tetrapeptides are replaced by Asn (see Supplementary Information S4), suggesting that ΔG_{app}^{flank} is indeed close to zero for the H-segments used to establish the ΔG_{app}^{aa} scale and that the true zero-point on the ΔG_{app}^{aa} scale is close to ΔG_{app}^{Ala}. Interestingly, studies of model peptides composed of 18–23 Ala residues have concluded that poly-Ala segments are at the threshold for transmembrane integration into pure lipid bilayers[9,19–21].

Effects of amino acid position on TM helix insertion

The ΔG_{app}^{aa} scale is strictly valid only for residues placed in the middle of the H-segment. To examine how ΔG_{app}^{aa} varies with residue position, we performed scans in which a pair of identical amino acid residues were moved symmetrically from the centre of the H-segment towards its amino and carboxy termini. Results for symmetrical scans with pairs of Leu, Phe, Trp and Tyr residues are shown in Fig. 3a. Leu and Phe behave similarly and show a small but significant position-specific variation of ± 0.2 kcal mol^{-1} around the mean. Trp and Tyr show markedly different behaviour: they strongly reduce membrane insertion when placed centrally, but become much less unfavourable as they are moved apart. Indeed, Trp is as favourable as Leu when placed in the outermost positions (see the construct 4L/15A in Fig. 3a). In membrane proteins, Trp and Tyr (in contrast to Phe and Leu) are over-represented near the ends of TM helices[11,12,22] and are known to interact preferentially with the headgroup region of model phospholipid bilayers[23–25]. The results presented here provide further support for the idea that protein–lipid interactions are central to the recognition of TM helices by the translocon.

Can we rationalize the variation in ΔG_{app} around the mean in the symmetrical scans? An obvious possibility is that these variations reflect the amphiphilicity of the H-segment. We used the ΔG_{app}^{aa} scale to compute the helical hydrophobic moment[26] for the central 19-residue segment in symmetrical scans with Leu and Ser residues, and then calculated the least-squares linear fit between experimentally-derived ΔG_{app} values and the hydrophobic moment. The ΔG_{app} values can be quite well reproduced by this procedure (Fig. 3b, c). The least amphiphilic H-segments (lowest hydrophobic moment) insert most efficiently into the membrane, and vice versa. The picture is similar for symmetrical scans with Lys and Asn residues, (Fig. 3d). Terminal Asn and Lys residues are much less detrimental to membrane insertion than centrally located ones, and there is a very strong amphiphilicity effect for H-segments with the two polar residues separated by six residues (that is, with the polar residues on the same face of an α-helix), which have greatly increased ΔG_{app} values. An interesting possibility suggested by these findings is that the H-segment forms an α-helix oriented so that the less hydrophobic face is in contact with the translocon protein and the more hydrophobic face is in contact with lipid during the membrane insertion process, consistent with crosslinking data[27,28].

To test further the importance of helical structure, scans with a single Pro or a pair of Pro residues were also done. In contrast with all the other residues tested, a pair of Pro residues maximally reduces the efficiency of membrane insertion at intermediate separations (4–10 residues; Fig. 3e). The positional dependence in ΔG_{app} is very strong, with a 1 kcal mol^{-1} difference between a centrally located Pro-Pro pair and a pair separated by 4–10 residues, and an even greater ΔG_{app} difference between centrally and terminally located Pro pairs. Furthermore, the scan with a single Pro residue (Fig. 3f) reveals that H-segments with a Pro residue in any of the three N-terminal positions of the hydrophobic stretch integrate much more efficiently than H-segments with Pro in the three C-terminal positions. In globular proteins, Pro is frequently found in the three N-terminal positions of α-helices, but is almost completely absent from the central and C-terminal parts[29,30]. Likewise, Pro is well tolerated near the N termini of TM helices in bacteriorhodopsin[31]. We conclude that helix formation is an important part of the membrane insertion process.

Can the biological ΔG_{app}^{aa} scale presented here be used to calculate the membrane insertion efficiency of natural polypeptide segments? It could in principle, but the scale is still too crude to be used for precise calculations of this kind because of the positional dependence of the ΔG_{app}^{aa} values, the contribution of charged flanking residues to ΔG_{app} (ref. 32), and the possible role of helix–helix interactions during insertion of multi-spanning membrane proteins[33]; further work will be necessary to quantify these effects fully.

In summary, the similarity between the biological ΔG_{app}^{aa} scale and both biophysical and statistical hydrophobicity scales, together with the positional preferences seen for residues such as Trp, Tyr, Ser, Asn, Lys and Pro, strongly suggest that direct protein–lipid interactions are essential for the recognition of TM helices by the translocon, and support models based on a partitioning of the TM helices between the Sec61 translocon and the surrounding lipid[5]. Exactly how this is managed is not apparent from the archaeal translocon structure, which is observed in a closed state without a translocating TM segment. Presumably, the open state is a highly dynamic one that permits rapid sampling of the translocon–bilayer interface by the translocating polypeptide. □

Methods

Enzymes and chemicals

All enzymes, the plasmid pGEM1, dithiothreitol (DTT) and the TnT coupled transcription/translation system were from Promega. ^{35}S-Met, ribonucleotides, deoxyribonucleotides, and dideoxyribonucleotides were from Amersham-Pharmacia. Oligonucleotides were obtained from Cybergene and MWG Biotech AG.

DNA manipulations

For expression of H-segment-containing Lep constructs from the pGEM1 plasmid, the 5′ end of the *lep* gene from *Escherichia coli* was modified by the introduction of an *XbaI* site and by changing the context 5′ of the initiator ATG codon to a Kozak consensus sequence[34]. Site-directed mutagenesis was used to introduce acceptor sites for N-linked glycosylation in positions 96–98 (Asn-Ser-Thr) and 258–260 (Asn-Ala-Thr).

Oligonucleotides encoding the different H-segments were introduced between a *SpeI* cleavage site in codons 226–227 and the *Kpn*I cleavage site in codon 253 of the *lep* gene[13]. H-segments were constructed using two or three double-stranded oligonucleotides (18–48 nucleotides long) with overlapping overhangs at the ends. Pairs of complementary oligonucleotides were first annealed at 85 °C for 10 min followed by slow cooling to 30 °C, after which the two or three annealed double-stranded oligos were mixed, incubated at 65 °C for 5 min, cooled slowly to room temperature and ligated into the vector. All H-segment inserts were confirmed by sequencing of plasmid DNA.

Expression *in vitro*

Constructs in pGEM1 were transcribed and translated in the Gold TnT Express 96 or TnT Quick systems (Promega). 1 μg DNA template, 1 μl ^{35}S-Met (5 μCi) and 1 μl microsomes (a gift from M. Sakaguchi) were added at the start of the reaction, and samples were incubated for 90 min at 30 °C. Translation products were analysed by SDS–PAGE and gels were quantified using a Fuji FLA-3000 phosphoimager and Image Reader 8.1j software. The membrane-insertion probability of a given H-segment was calculated as the quotient between the intensity of the singly glycosylated band divided by the summed intensities of the singly glycosylated and doubly glycosylated bands. Mean values from at least four independent experiments were used in the derivation of the ΔG_{app}^{aa} scale, and from at least two independent experiments for the other results reported. On average, glycosylation levels vary by no more than ±2% between repeat experiments, corresponding to a standard deviation of ± 0.1 kcal mol^{-1} in the ΔG_{app} values.

Expression *in vivo*

Protein synthesis in BHK cells using the Semliki forest virus (SFV) expression system was carried out as described previously[35,36]. Briefly, Lep constructs under the SP6 promoter in the SFV vector were linearized for *in vitro* transcription. The resulting RNA was used to transfect BHK cells by electroporation. Six hours after electroporation, cells were starved of Met for 30 min, then labelled with ^{35}S-Met for 15 min. Cells were solubilized in lysate buffer containing 1% nonidet P-40 and protease inhibitors, immunoprecipitated using a Lep antiserum and analysed by SDS–PAGE. Quantifications were carried out on a phosphorimager as described above.

Hydrophobic moment calculation

Hydrophobic moments (μ) were computed using MPEx (http://blanco.biomol.uci.edu/mpex/) and the ΔG_{app}^{aa} scale. Linear fits ($\Delta G_{app} = a_i\mu + h_i$) were calculated between μ and the ΔG_{app} values for a given symmetrical pair scan (i), excluding the data point representing the maximum separation (17 residues) between the scanned residues. For the

2L/17A series, the point representing a separation of 3 residues was also excluded. Finally, the a_i values were averaged for the pair scans with 2L/17A, 3L/16A, 4L/15A and 2S/4L/13A, yielding the mean value $a = 0.17$ (standard deviation, ± 0.04).

Received 31 August; accepted 16 November 2004; doi:10.1038/nature03216.

1. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model. Application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
2. von Heijne, G. Recent advances in the understanding of membrane protein assembly and structure. *Q. Rev. Biophys.* **32,** 285–307 (2000).
3. von Heijne, G. Membrane protein assembly *in vivo*. *Adv. Protein Chem.* **63,** 1–18 (2003).
4. Snapp, E., Reinhart, G., Bogert, B., Lippincott-Schwartz, J. & Hegde, R. The organization of engaged and quiescent translocons in the endoplasmic reticulum of mammalian cells. *J. Cell Biol.* **164,** 997–1007 (2004).
5. Rapoport, T. A., Goder, V., Heinrich, S. U. & Matlack, K. E. Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* **14,** 568–575 (2004).
6. Alder, N. N. & Johnson, A. E. Cotranslational membrane protein biogenesis at the endoplasmic reticulum. *J. Biol. Chem.* **279,** 22787–22790 (2004).
7. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427,** 36–44 (2004).
8. Woolhead, C. A., McCormick, P. J. & Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding. *Cell* **116,** 725–736 (2004).
9. de Planque, M. R. R. & Killian, J. A. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. *Mol. Membr. Biol.* **20,** 271–284 (2003).
10. White, S. H. & Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 319–365 (1999).
11. Ulmschneider, M. B. & Sansom, M. S. P. Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta* **1512,** 1–14 (2001).
12. Beuming, T. & Weinstein, H. A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. *Bioinformatics* **20,** 1822–1835 (2004).
13. Sääf, A., Wallin, E. & von Heijne, G. Stop-transfer function of pseudo-random amino acid segments during translocation across prokaryotic and eukaryotic membranes. *Eur. J. Biochem.* **251,** 821–829 (1998).
14. Wimley, W. C., Creamer, T. P. & White, S. H. Solvation energies of amino acid sidechains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35,** 5109–5124 (1996).
15. Jayasinghe, S., Hristova, K. & White, S. H. Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **312,** 927–934 (2001).
16. Cornette, J. L. *et al.* Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* **195,** 659–685 (1987).
17. Degli Esposti, M., Crimi, M. & Venturoli, G. A critical evaluation of the hydropathy profile of membrane proteins. *Eur. J. Biochem.* **190,** 207–219 (1990).
18. Heinrich, S., Mothes, W., Brunner, J. & Rapoport, T. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102,** 233–244 (2000).
19. Lu, L. P. & Deber, C. M. Guidelines for membrane protein engineering derived from de novo designed model peptides. *Biopolymers* **47,** 41–62 (1998).
20. Bechinger, B. Membrane insertion and orientation of polyalanine peptides: A N-15 solid-state NMR spectroscopy investigation. *Biophys. J.* **81,** 2251–2256 (2001).
21. Lewis, R. N. *et al.* A polyalanine-based peptide cannot form a stable transmembrane alpha-helix in fully hydrated phospholipid bilayers. *Biochemistry* **40,** 12103–12111 (2001).
22. Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G. & Elofsson, A. Architecture of helix bundle membrane proteins: An analysis of cytochrome c oxidase from bovine mitochondria. *Protein Sci.* **6,** 808–815 (1997).
23. Killian, J. A. & von Heijne, G. How proteins adapt to a membrane-water interface. *Trends Biochem. Sci.* **25,** 429–434 (2000).
24. Yau, W. M., Wimley, W. C., Gawrisch, K. & White, S. H. The preference of tryptophan for membrane interfaces. *Biochemistry* **37,** 14713–14718 (1998).
25. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Struct. Biol.* **3,** 842–848 (1996).
26. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* **179,** 125–142 (1984).
27. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J. & Rapoport, T. A. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94,** 795–807 (1998).
28. McCormick, P. J., Miao, Y., Shao, Y., Lin, J. & Johnson, A. E. Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. *Mol. Cell* **12,** 329–341 (2003).
29. Presta, L. G. & Rose, G. D. Helix signals in proteins. *Science* **240,** 1632–1641 (1988).
30. Richardson, J. S. & Richardson, D. C. Amino acid preferences for specific locations at the ends of α-helices. *Science* **240,** 1648–1652 (1988).
31. Yohannan, S. *et al.* Proline substitutions are not easily accommodated in a membrane protein. *J. Mol. Biol.* **341,** 1–6 (2004).
32. Kuroiwa, T., Sakaguchi, M., Mihara, K. & Omura, T. Systematic analysis of stop-transfer sequence for microsomal membrane. *J. Biol. Chem.* **266,** 9251–9255 (1991).
33. Anthony, V. & Skach, W. R. Molecular mechanism of P-glycoprotein assembly into cellular membranes. *Curr. Protein Pept. Sci.* **3,** 485–501 (2002).
34. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234,** 187–208 (1999).
35. Liljeström, P. & Garoff, H. A new generation of animal cell expression vectors based on the Semliki Forest virus replicon. *Biotechnology* **9,** 1356–1361 (1991).
36. Liljeström, P., Lusa, S., Huylebroeck, D. & Garoff, H. *In vitro* mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *J. Virol.* **65,** 4107–4113 (1991).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We wish to thank E. Missioux for technical assistance and R. MacKinnon, D. Rees, and T. Rapoport for comments. This work was supported by grants from the Swedish Cancer Foundation to G.v.H. and I.M.N., the Marianne and Marcus Wallenberg Foundation and the Swedish Research Council to G.v.H., the Magnus Bergvall Foundation to I.M.N., and the National Institute of General Medical Sciences to S.H.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.v.H. (gunnar@dbb.su.se).