doi:10.1016/j.jmb.2009.11.036 J. Mol. Biol. (2010) 396, 221–229

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context

Linnea E. Hedin[1]†, Karin Öjemalm[1]†, Andreas Bernsel[1,2], Aron Hennerdal[1,2], Kristoffer Illergård[1,2], Karl Enquist[1], Anni Kauko[1,2], Susana Cristobal[1], Gunnar von Heijne[1,2], Mirjam Lerch-Bader[1]*, IngMarie Nilsson[1]* and Arne Elofsson[1,2]*

[1]*Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden*

[2]*Stockholm Bioinformatics Center, AlbaNova University Centre, SE-106 91 Stockholm, Sweden*

Received 23 September 2009; received in revised form 11 November 2009; accepted 13 November 2009
Available online 18 November 2009

Edited by J. Bowie

In mammalian cells, most integral membrane proteins are initially inserted into the endoplasmic reticulum membrane by the so-called Sec61 translocon. However, recent predictions suggest that many transmembrane helices (TMHs) in multispanning membrane proteins are not sufficiently hydrophobic to be recognized as such by the translocon. In this study, we have screened 16 marginally hydrophobic TMHs from membrane proteins of known three-dimensional structure. Indeed, most of these TMHs do not insert efficiently into the endoplasmic reticulum membrane by themselves. To test if loops or TMHs immediately upstream or downstream of a marginally hydrophobic helix might influence the insertion efficiency, insertion of marginally hydrophobic helices was also studied in the presence of their neighboring loops and helices. The results show that flanking loops and nearest-neighbor TMHs are sufficient to ensure the insertion of many marginally hydrophobic helices. However, for at least two of the marginally hydrophobic helices, the local interactions are not enough, indicating that post-insertional rearrangements are involved in the folding of these proteins.

Keywords: marginally hydrophobic helix; membrane protein; topology; hydrophobicity

Introduction

While it is generally recognized that hydrophobicity is the overriding characteristic of transmembrane α-helices (TMHs) in integral membrane proteins, all attempts to define a simple "threshold hydrophobicity" that determines whether or not a given polypeptide segment in a multispanning membrane protein will form a TMH have failed. In contrast, such a threshold appears to exist for TMHs in single-spanning membrane proteins, as has been shown recently.[1] This suggests that some TMHs in multispanning proteins may depend on other parts of the same protein for efficient insertion and folding.[2] Indeed, a small number of such "marginally hydrophobic" transmembrane helices (mTMHs) that do not insert into the membrane by themselves have been identified in P-glycoprotein, the cystic fibrosis transmembrane conductance regulator,[3] aquaporin-1 (AQP1),[4] and the plant K_v channel KAT1.[5]

How common are mTMHs, and how much of the surrounding parts of the protein are needed for them to insert efficiently across the membrane? Here, we have addressed these questions systematically by first identifying candidate mTMHs in multispanning proteins of known three-dimensional structure using the "ΔG predictor,"[1] and then experimentally measuring their apparent free energy of insertion (ΔG_{app}) into the endoplasmic reticulum (ER) membrane, both with and without inclusion of their immediate flanking loop segments. In agreement with the predictions, we find that most of the candidate mTMHs indeed do not

*Corresponding authors. A. Elofsson is to be contacted at Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail addresses: mirjam@dbb.su.se; ingmarie@dbb.su.se; arne@bioinfo.se.

† L.E.H. and K.Ö. contributed equally to this work.

Abbreviations used: TMH, transmembrane helix; mTMH, marginally hydrophobic transmembrane helix; ER, endoplasmic reticulum; RM, dog pancreas rough microsomes; Lep, leader peptidase.

insert into the ER membrane by themselves. The inclusion of upstream and downstream flanking loops improves membrane insertion for some mTMHs. Finally, for five mTMHs that do not insert even when flanking loops are included, we studied insertion in the presence of one or both flanking TMHs. For three of these five mTMHs, inclusion of neighboring TMHs leads to efficient insertion into the membrane. These results suggest that the insertion propensity of mTMHs in most, but not all, cases depends on "local" sequence characteristics (i.e., on the flanking loops and the nearest-neighbor TMHs).

Results

Selection of mTMHs and experimental setup

In order to focus on the most extreme cases, we define mTMHs as those TMHs in the known three-dimensional membrane protein structures for which the ΔG predictor predicts ΔG_{app}>1.4 kcal/mol. Twenty percent of the 388 TMHs found in a nonredundant set of 102 α-helical membrane proteins with known structure (see Materials and Methods) were classified as mTMHs by this criterion, and we chose 16 of these, all with their N-terminal ends in the lumen (N_{out}), for further studies. N_{in}-helices were excluded here because it has been shown that the ΔG predictor results need to be slightly adjusted to accurately predict their insertion efficiencies.[6]

To experimentally measure ΔG_{app} for the mTMHs, we used a previously described assay that allows us to determine the efficiency of insertion into ER-derived rough microsomal membranes.[6,7] Briefly, the mTMH segment of interest is introduced into the "host" protein leader peptidase (Lep) (Fig. 1a). Lep contains two N-terminal TMHs and a large C-terminal domain (P2). When expressed in a rabbit reticulocyte *in vitro* transcription/translation system in the presence of dog pancreas rough microsomes (RMs), Lep inserts into the microsomal membrane with both the short N-terminal tail and the large P2 domain located in the lumen of the microsomes.[8] The presence of Lep TMH1 ensures SRP-dependent targeting of all constructs to the Sec61 translocon.[9,10]

In the Lep construct, the TMH segment to be tested is placed near the middle of the P2 domain and is flanked by two engineered Asn-X-Thr acceptor sites for N-linked glycosylation (G1 and G2). If the TMH segment is inserted into the membrane, only the G1 site will be modified by the lumenally orientated oligosaccharyl transferase; if the TMH segment is translocated across the membrane, both the G1 and G2 sites will receive a glycan. The fraction of singly (f_{1x}) and doubly (f_{2x}) glycosylated protein can be quantified by phosphorimager analysis of SDS-PAGE gels, and the apparent change in free energy upon insertion can be calculated using the relation $\Delta G_{app} = -RT \ln K$, where R is the gas constant, T the absolute temperature, and $K = \frac{f_{1x}}{f_{2x}}$. "Insulating" tetrapeptides (GGPG...GPGG) were placed on either side of all TMH segments in order to minimize the influence of the surrounding Lep sequence.[7] Since very low and very high insertion efficiencies cannot be accurately measured, ΔG_{app} values outside the interval ±1.5 kcal/mol are only qualitative. While many of the mTMHs tested here come from prokaryotic proteins, the high degree of sequence conservation in the Sec-translocon components between prokaryotes and eukaryotes[11] suggests that general conclusions can be drawn based on studies using the microsomal assay system. Our observation that a topology prediction algorithm based on measurements carried out using this system works well for both prokaryotic and eukaryotic membrane proteins[12] provides further support for this contention.

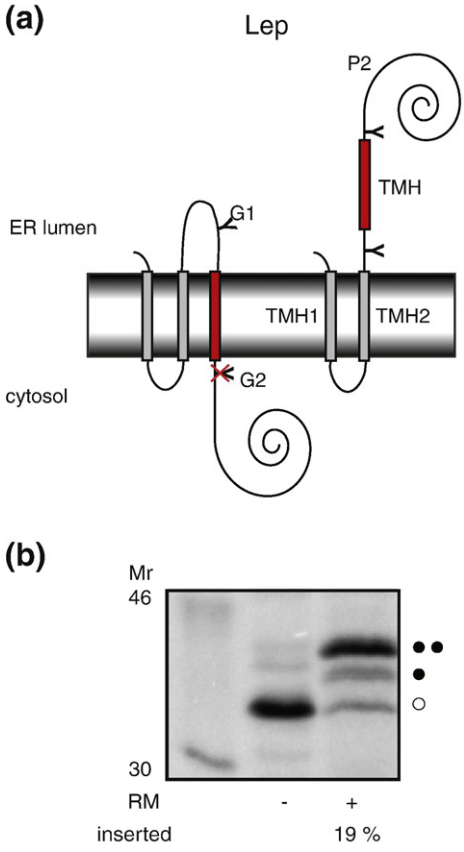

Fig. 1. Membrane integration assay. (a) Lep was used as model protein.[6] The engineered TMH segment of interest (red bar) is inserted into Lep as shown. The insertion of the TMH segment across the dog pancreas rough microsomal membrane during *in vitro* expression gives a protein glycosylated only on the G1 site. (b) Representative SDS-PAGE gel for a Lep construct (mTMH=CyoB TMH6). The construct was expressed in rabbit reticulocyte lysate in the presence (+RM) or absence (−RM) of dog pancreas RMs. Unglycosylated (○), singly glycosylated (l), and doubly (ll) glycosylated forms are indicated.

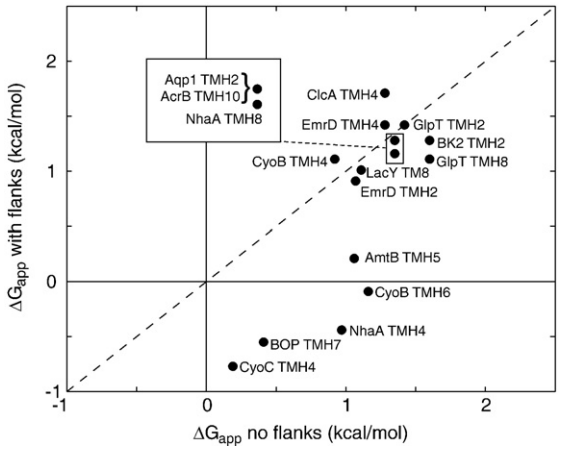

Fig. 2. Free energy of insertion for mTMHs with and without flanking loops. The name of the protein and the number of TMH are shown. Five helices insert significantly better when their flanks are included: CyoB-TMH4, CyoC-TMH4, AmtB-TMH5, NhaA-TMH4, and BOP-TMH7. For these five mTMHs, the biological hydrophobicity scale predicts segments of lower hydrophobicity when the flanks are included and the predictor is set to allow longer transmembrane segments (up to 30 residues) (see Supplementary Table S1). In addition, the positive-inside bias is generally strengthened by the added sequences. For the one case where the recognition of the mTMH by the translocon decreases (EmrD TMH2), the added flanks weaken the positive-inside bias. Note that two mTMHs gave identical values both with and without flanks (Aqp1 TMH2 and AcrB TMH10).

Membrane insertion of individual mTMHs

All 16 mTMHs in our data set have a predicted ΔG_{app}>1.4 kcal/mol and have their N-terminal ends in the lumen or on the extracellular (periplasmic) side in the native protein structure (N_{out}–C_{in} topology). Eleven of the 16 mTMHs insert poorly when tested in the Lep construct, while four mTMHs insert unexpectedly well (ΔG_{app}<1 kcal/mol): CyoC TMH4, CyoB TMH4, NhaA TMH4, and BOP TMH7 (Supplementary Table S1). These four helices have the lowest predicted ΔG_{app} of the investigated mTMHs (1.4–1.8 kcal/mol). As an example, ΔG_{app} for CyoC TMH4 is 0.2 kcal/mol in our experiments compared to the predicted 1.4 kcal/mol. CyoC TMH4 contains several charged residues towards the cytosolic end of the TMH. We have shown previously that positive charges on the cytosolic side of a TMH can aid insertion,[13] as expected from the so-called "positive-inside" rule.[14] Also, molecular simulation studies show that positively charged residues can lower their ΔG_{app} by interacting with lipid phosphate and carbonyl groups as well as by pulling water molecules into the membrane.[15,16] BOP TMH7 (predicted ΔG_{app}=1.4 kcal/mol, measured ΔG_{app}=0.4 kcal/mol) has two positively charged residues near the cytosolic end that may contribute to the low observed ΔG_{app}. Previous tests have shown that predicted ΔG_{app} values generally are within ±0.5 kcal/mol of the measured values,[1] but given the simplicity of the additive model used to predict ΔG_{app} values and the fact that it does not take flanking charged residues into account, it is perhaps not surprising that a few mTMHs insert better than predicted.

Flanking loops can favor insertion of marginally hydrophobic helices

We have found previously that loop segments flanking a TMH can affect its insertion into the membrane.[13] We therefore added the adjacent N- and C-terminal loops to the mTMHs and remeasured ΔG_{app}. Predicted ΔG_{app} was recalculated allowing for a larger sequence window. In case of loops longer than 20 residues, only the 20 residues flanking the mTMH were included.

As seen in Fig. 2, inclusion of the flanking loops clearly favors insertion for 5 of the 16 mTMHs. Strong reduction in ΔG_{app} is seen for CyoB TMH6 (from 1.2 to −0.1 kcal/mol), CyoC TMH4 (from 0.2 to −1.8 kcal/mol), AmtB TMH5 (from 1.1 to 0.2 kcal/mol), NhaA TMH4 (from 1.0 to −0.4 kcal/mol), and BOP TMH7 (from 0.4 to −0.6 kcal/mol). None of the most inefficiently inserting mTMHs (measured $\Delta G_{app} \geq 1.2$ kcal/mol without flanks) insert markedly better when the flanking loops are added, while five of seven with lower ΔG_{app} do. A more detailed analysis of these flanking sequences showed that when allowing for longer transmembrane segment, the predicted ΔG_{app} dropped significantly for all these five sequences. In four cases, the free energy of insertion dropped to below 1 kcal/mol (see Supplementary Table S1). This shows that sequence features outside the TMH itself can improve membrane insertion, at least for mTMHs with a hydrophobicity not too far from the threshold. For most of these helices, the number of positively charged residues in cytosolic loops increases when the flanks are included, indicating that the positive-inside rule may be one important factor contributing to the increased insertion efficiency. In one case (EmrD TMH2), inclusion of the flanking loops decreased the insertion efficiency from 8.5% to 0.5%. Although the assay used is not particularly sensitive for such small insertion efficiencies, it can be noted that this is one of only two cases where the inclusion of flanks adds more positively charged residues to the lumenal loop, violating the positive-inside rule.

Membrane insertion of mTMHs in the presence of flanking transmembrane helices

It has been shown previously that in some cases, a neighboring TMH can favor the membrane insertion of a mTMH[17] and that there is a correlation between the polarity of a TMH and its interaction area with the rest of the protein.[12] The comprehensive collection of mTMHs studied here is, on average, found to be more buried than other TMHs ($p<2\times10^{-6}$ as judged by a two-sided t test; data not shown). However, a large part of their buried

1 mTMH	2 mTMH	3 TMH_{pre}	4 TMH_{sub}		5 TMH_{pre} + mTMH		6 mTMH + TMH_{sub}		7 mTMH + TMH_{sub}		8 TMH_{pre} + mTMH + TMH_{sub}	
Topology	(a)	(b)	(c)	(d)	(e) (f)	(g) (h)	(i) (j)	(k) (l)	(m) (n)	(o) (p)	(q) (u)	(r) (v)	(s) (w)	(t) (x)
AcrB TMH10	10 %	96 %	100 %	70 %	35 %	65 %	I + J 54 %	K + L 46 %	M 18 % N 61 %	O 7 % P 14 %	R 8 % U 14 %		T 78 % W 0 %	
AQP1 TMH 2	10 %	100 %	100%	70 %	12 %	88 %	I + J 73 %	K + L 27 %	M 19 % N 75 %	O 2 % P 4 %	R 14 % U 82 %		T 4 % W 0 %	
EmrD TMH 2	1 %	97 %	100 %	81 %	10 %	90 %	I + J 45 %	K + L 55 %	M + N 46 %	O 43 % P 11 %	R + U 42 %		T 58 % W 0 %	
EmrD TMH 4	8 %	100 %	83 %	16 %	5 %	95 %	I + J 17 %	K + L 83 %	M + N 28 %	O 5 % P 67 %	R + U 24 %		T 7 % W 69 %	
GlpT TMH 8	13 %	82 %	100 %	90 %	15 %	85 %	I + J 34 %	K + L 66 %	M + N 45 %	O 55 % P 0 %	R + U 19 %		T 81 % W 0 %	

Fig. 3. Effect of neighboring TMHs on the insertion of mTMHs. Red bar, mTMH; yellow bar, preceding TMH (TMH_{pre}); blue bar, subsequent TMH (TMH_{sub}); grey bars, TMHs from Lep and Lep′. All glycosylation sites are denoted by Y. The five investigated mTMHs are given in column 1. Columns 2–4 give the insertion efficiencies for mTMH, TMH_{pre} and TMH_{sub} on their own. For TMH_{sub}, the propensity to insert with the wrong topology was also investigated (column 4, topology D). For columns 5–8, the investigated protein construct is shown on top; note the use of Lep (mTMH + TMH_{sub}) or Lep′ (TMH_{pre} + mTMH and TMH_{pre} + mTMH + TMH_{sub}). The result for each construct is presented as the fraction (in percent) of the glycosylated protein species that adopts the topology shown straight above in the row "Topology," with a Y denoting here a glycosylation acceptor site that has received a glycan. Figures of and results for topologies that can be distinguished are separated by dashed lines. Since the insertion of all TMH_{pre} segments was very efficient (see column 3), all theoretically possible topologies in the more complex constructs where a TMH_{pre} segment does not span the membrane were considered improbable, indicated by these topologies being crossed out and ignored in the interpretation of the results (topologies F and H in column 5 and topologies Q, S, V, and X in column 8). Note that topologies I and J, K and L in column 6 cannot be distinguished and the given results are for I + J and K + L, respectively. Note also that columns 6 and 7 both show mTMH + TMH_{sub} in Lep′, but with two and four glycosylation sites, respectively. For EmrD TMH2, EmrD TMH4, and GlpT TMH8, the loop after the mTMH is too short to be glycosylated, and thus topologies M and N, R and U cannot be distinguished for these constructs.

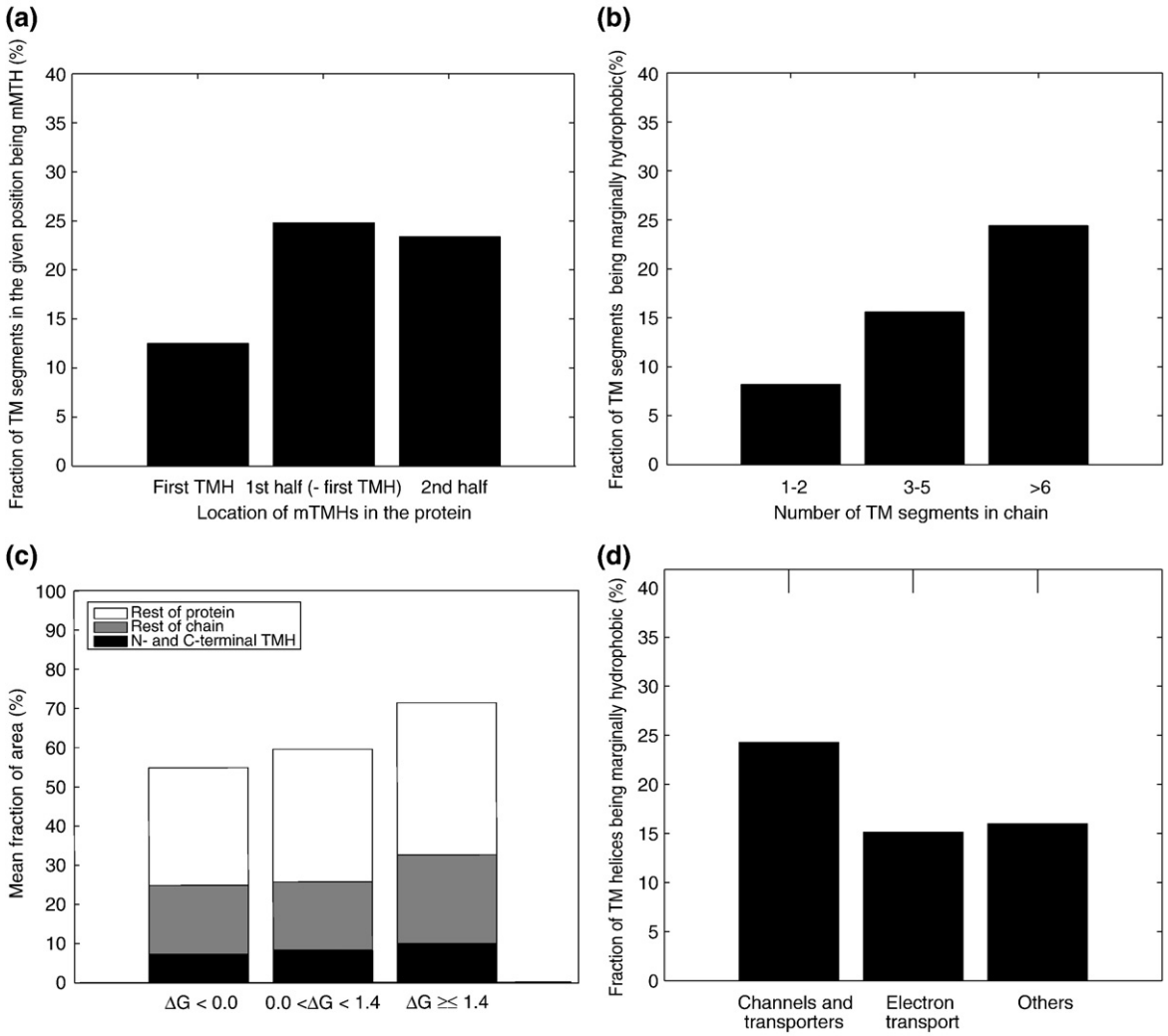

Fig. 4. Statistical analysis of 388 TMHs from 102 α-helical transmembrane protein chains with known three-dimensional structure. (a) Location of strongly (ΔG_{app}<0.0 kcal/mol; black), moderately (0.0 kcal/mol ≤ ΔG_{app}< 1.4 kcal/mol; gray), or marginally ($\Delta G_{app} \geq 1.4$ kcal/mol; white) hydrophobic helices within the protein chain. (b) Fraction of TMHs being predicted to be strongly, moderately, or marginally hydrophobic, grouped according to the number of TMHs in the protein chain. (c) Fraction surface area buried against TMH_{pre} and TMH_{sub} (black), against other parts of the polypeptide chain (gray), and against other subunits (for oligomeric proteins; white) for TMHs of different hydrophobicity. (d) Fraction of TMHs predicted to be strongly, moderately, or marginally hydrophobic for ion channels and small-molecule transporters, proteins involved in electron transport, and other types of membrane proteins.

surface area involves other chains in the same complex, and it is unlikely that such subunit–subunit interactions are formed already during the translocon-mediated membrane insertion step of a given subunit.

To investigate to what degree the insertion of mTMHs can be affected by the presence of neighboring TMHs, we focused on those mTMHs that have high experimental ΔG_{app} values even when the flanking loops are included. To facilitate the interpretation of the experimental results, we further required that the selected mTMHs were flanked by TMHs that insert efficiently on their own (measured ΔG_{app}<<0 kcal/mol; see Supplementary Table S2). Since these TMHs have an N_{in}–C_{out} orientation in the membrane, another Lep construct (Lep′) was used to measure their membrane insertion propensity. In this Lep′ construct (Supplementary Fig. S1), the TMH segment replaces the Lep TMH2. The glycosylation acceptor site (G2′) located in the beginning of the P2 domain will be modified only if the TMH segment inserts into the membrane, while the G1′ site, embedded in an extended N-terminal sequence of 24 amino acids, is always glycosylated. Hence, the insertion efficiency for single-spanning TMH segments introduced in Lep′ constructs is given by the fraction of doubly glycosylated protein (f_{2x}). To facilitate the interpretation of glycosylation patterns from multi-helix constructs, the propensity of the mTMH-flanking succeeding TMH to insert in its reverse, N_{out}–C_{in}, orientation was also measured using the original Lep construct. We found five mTMHs that fulfilled all requirements, namely, AcrB TMH10, AQP1

TMH2, EmrD TMH2, EmrD TMH4, and GlpT TMH8 (Fig. 3 and Supplementary Table S2). The insertion of AQP1 TMH2 has been studied in detail before,[18] and we included it here as a control.

Each selected mTMH was analyzed as part of a longer segment including the preceding TMH (TMH_{pre}), the subsequent TMH (TMH_{sub}), or both, as well as the loops upstream of TMH_{pre} and downstream of TMH_{sub} (up to a maximum length of 20 amino acids).

In summary, segments consisting of mTMH + TMH_{sub} or just TMH_{sub} (to test its insertion when in the reverse orientation) were introduced into the standard Lep construct. Segments containing TMH_{pre}, TMH_{sub}, TMH_{pre} + mTMH, or TMH_{pre} + mTMH + TMH_{sub} were introduced into the Lep′ construct. The membrane insertion efficiency of each construct was then measured as previously described (also see Materials and Methods). The results are given in Supplementary Table S2 and summarized in Fig. 3.

For Lep variants with more than one guest TMH, different theoretically possible topological forms can give rise to identical glycosylation patterns. We therefore engineered additional glycosylation sites into the loop after the mTMH and into the C-terminal tail of the multi-helix Lep′ constructs (Supplementary Fig. S2 and Supplementary Tables S2 and S3). This allowed us to achieve easily interpretable patterns for constructs containing mTMH + TMH_{sub}. However, for EmrD TMH2, EmrD TMH4, and GlpT TMH8, the loop after the mTMH is too short to be glycosylated. For these mTMHs, it is therefore not possible to distinguish between topologies M and N for the TMH_{sub} constructs (Fig. 3). For the constructs containing TMH_{pre} + mTMH + TMH_{sub}, the interpretation is still straightforward, since we know that both the preceding and the subsequent TMH insert efficiently by themselves. We also know the propensity of the TMH_{sub} to insert with the reverse topology, and we can further confirm our interpretations by using a protease protection assay (see Materials and Methods and Supplementary Fig. S3).

The nearest-neighbor TMHs can influence the membrane insertion of a mTMH

As seen in Fig. 3, AcrB TMH9 improves the insertion of the AcrB TMH10 to some extent (from 10% to 35%), but none of the other preceding helices improves the insertion of the neighboring mTMH. The subsequent TMH significantly favors the insertion of the neighboring mTMH in EmrD TMH2 (from 1% to 43%) and GlpT TMH8 (from 13% to 55%). Finally, the correct topology is attained when both the previous and the subsequent TMHs are present for AcrB TMH10 (78%), EmrD TMH2 (58%), and GlpT TMH8 (81%), but not for EmrD TMH4 or AQP1 TMH2.

In cases where the subsequent TMH inserts efficiently only in its natural but not in its reverse orientation, an inefficiently inserting mTMH can destroy the insertion of the downstream TMH as well, as illustrated by EmrD TMH4 where the inefficient insertion of TMH4 leads to an inverted, less well-inserting orientation of TMH5 and the incorrect topologies L and W (Fig. 3). We could confirm this observation using a protease protection assay (Supplementary Fig. S3). Our results for AQP1 TMH2 agree with previous data showing that this protein attains its final six-TMH topology only after a large rearrangement of an intermediate four-TMH form in which TMH2 and TMH4 do not span the membrane.[18]

Discussion

Transmembrane α-helices of low hydrophobicity (mTMHs) are quite frequent in multispanning membrane proteins. They often contain functionally important polar or charged residues that render their presence within the lipid bilayer energetically unfavorable. They more often contain non-helical elements than more hydrophobic helices ($p<0.03$ by Fisher's exact test; data not shown). They are rarely found as the N-terminal TMH in proteins (Fig. 4A) (possibly because the N-terminal TMH often serves as the targeting sequence recognized by the signal recognition particle[9]); their frequency increases with the number of TMHs in the protein (Fig. 4B); and they tend to have less lipid-exposed surface area than more hydrophobic TMHs (Fig. 4C). In fact, 38 of the 52 polar residues (73%) in our collection of mTMHs are buried in the three-dimensional structure. Of the remaining 14 non-buried ones, 11 (78%) are located more than 9 Å away from the center of the membrane and may reach into the lipid headgroup region. Not surprisingly, given their ability to transport hydrophilic compounds, ion channels and small-molecule transporters seem to contain a higher fraction of mTMHs than other membrane proteins (Fig. 4D) ($p<0.03$ by Fisher's exact test). Understanding how mTMHs can be inserted into and maintained within biological membranes is of importance not only for elucidating the process of insertion itself but also for membrane protein topology prediction and for the growing field of membrane protein engineering.

In this work, we show that 11 of the 16 TMHs identified by the ΔG predictor indeed cannot insert by themselves into the ER membrane, while the remaining five insert rather inefficiently. The ΔG predictor is thus a good guide to assess the "intrinsic" insertion efficiency of individual TMHs. Further, we show that the insertion efficiency of most mTMHs can be substantially increased both by the inclusion of flanking loop segments and by the presence of well-inserting neighboring N- and C-terminal TMHs. For most mTMHs, local sequence context (up to nearest-neighbor TMHs) thus seems to provide sufficient guidance for proper membrane insertion, although, in addition to the previously described AQP1 TMH2,[4] we have uncovered one new case—EmrD TMH4—where this does not seem

to hold. More focused studies of the individual mTMHs identified here can now be carried out to pinpoint the precise molecular basis for the increased insertion efficiency in each case.

The relatively large fraction of TMHs that cannot insert efficiently into the membrane in the absence of other parts of the protein raises the interesting issue of how the Sec61 translocon in the ER can handle multiple TMHs, some of which are not sufficiently hydrophobic for membrane integration on their own, at the same time. Conceivably, two or possibly three TMHs can fit simultaneously into the translocon channel, provided that the hypothesized "lateral gate" in the channel wall[19] is open towards the surrounding lipid. If a sufficiently hydrophobic two- or three-helix bundle can form within the channel or possibly in the gate region, the whole assembly may be able to partition into the bilayer *en bloc* despite the presence of marginally hydrophobic TMHs. Further rearrangements of the TMHs may then lead to the final structure of the full-length protein.

Materials and Methods

Enzymes and chemicals

Unless stated otherwise, all chemicals were from Sigma-Aldrich (St. Louis, MO). Oligonucleotides were obtained from Sigma-Aldrich, Cybergene AB (Stockholm, Sweden) and MWG Biotech AG (Ebersberg, Germany). All enzymes were from Fermentas (Burlington, Ontario, Canada), except Phusion DNA polymerase from Finnzymes OY (Espoo, Finland) and QuikChange™ Site-Directed Mutagenesis Kit from Stratagene (La Jolla, CA). The plasmid pGEM-1 and the TNT® SP6 Quick Coupled Transcription/Translation System were from Promega Biotech AB (Madison, WI). [^{35}S]Met was from PerkinElmer (Boston, MA).

DNA manipulation

Double-stranded oligonucleotides encoding the different protein segments were introduced into the *lepB* gene as amplified PCR fragments using primers complementary to the 5′ and 3′ ends of the selected part of the gene. Fragments were amplified using Phusion DNA polymerase (Finnzymes OY) and cloned into pGEM1 containing the Lep constructs as a SpeI–KpnI fragment.[6,7] Glycosylation acceptor sites (NX[S/T]) found in the segments were mutated to QX[S/T], and extra glycosylation acceptor sites in multi-TMH segments were introduced using the QuikChange™ Site-Directed Mutagenesis Kit. The amplified DNA products were purified using the QIAquick PCR Purification Kit from QIAGEN (Hilden, Germany). All inserted segments were confirmed by sequencing of plasmid DNA at Eurofins MWG Operon (Ebersberg, Germany) or BM labbet AB (Furulund, Sweden).

Expression *in vitro*

Constructs in pGEM1 were transcribed and translated in the TNT® SP6 Quick Coupled System from Promega. DNA template (150–200 ng), 1 μl of [^{35}S]Met (15 μCi), and 1 μl of dog pancreas RMs were added to 10 μl of lysate at the start of the reaction, and samples were incubated for 90 min at 30 °C.[7]

Translation products were analyzed by SDS-PAGE, and proteins were visualized in a Fuji FLA-3000 phosphorimager using the Image Reader V1.8J/Image Gauge V 3.45 software. The MultiGauge (Fujifilm) software was used to generate a one-dimensional intensity profile of each gel lane, and the multi-Gaussian fit program from the Qtiplot software package‡ was used to calculate the peak areas of the glycosylated protein bands in the profile. The membrane insertion efficiency of a given guest segment was calculated as the quotient between the peak area of the singly glycosylated protein band and the summed peak areas of the singly glycosylated and doubly glycosylated protein bands. For segments with multiple TMHs and four glycosylation acceptor sites, the protein fraction of a particular topology, with n glycans, was calculated as the peak area of the n times glycosylated protein band divided by the summed peak areas of all glycosylated protein bands. On average, the glycosylation levels vary by no more than ±5% between repeated experiments, corresponding to an SD of ±0.25 kcal/mol in the ΔG_{app} values. All values were calculated as mean values from at least two independent experiments.

In those cases where an introduced segment replaces the second TMH in Lep (see Supplementary Table S2), a fraction of the molecules is often cleaved by signal peptidase to a smaller species in the presence of RMs.[6] The signal peptidase active site is located on the lumenal surface of the ER membrane, and the processed form thus originates from molecules in which the TMH segment is inserted into the membrane and the P2 domain is in the lumen. In the calculation of the insertion efficiency for the Lep′ constructs, the fraction of cleaved molecules is included in f_{2x} (see the main text). In case of Lep′ constructs with three TMH inserts, two cleaved fragments can be found. The shorter fragment is included in f_{3x} and the longer fragment in f_{2x}.

Proteinase K protection assay

Protein expression was carried out as described in "Expression *in vitro*." Samples were then placed on ice, and 1 μl of $CaCl_2$ solution (200 mM) and 0.5 μl of Proteinase K (4.5 U/μl) were added. After 30 min of incubation on ice, Proteinase K was inactivated with 2 μl of PMSF (20 mM ethanolic solution) and samples were centrifuged for 10 min at 80,000 rpm in a Beckman TL-100 centrifuge. Membrane pellets were resuspended in 50 μl of buffer (250 mM sucrose, 50 mM TEA, 1 mM DTT) and 2 μl of PMSF (20 mM ethanolic solution) and centrifuged for 10 min at 80,000 rpm. Pellets were resuspended in 10 μl of buffer (250 mM sucrose, 50 mM TEA, 1 mM DTT), 2 μl of PMSF (20 mM ethanolic solution), and 40 μl of sample buffer and analyzed by SDS-PAGE.

Bioinformatics

In the statistical analysis, a data set with 388 TMHs from 102 representative α-helical transmembrane protein chains from the Protein Data Bank[20] was used. The filtering procedure, including removal of badly resolved structures, identifying representatives at OPM superfamily level,[21]

‡ http://www.qtiplot.ro

and subsequent homology reduction, is described in detail elsewhere (A.K. *et al.*, unpublished results).

Predicted ΔG_{app} values were calculated using the ΔG predictor.[1] In short, the position-dependent contributions to the overall insertion free energy for each amino acid are summed across the sequence of a TMH, after which corrections for segment length and hydrophobic moment are applied. In the calculation of ΔG_{app} for the mTMH sequences without flanks, the "allow subsequence" option was turned on. For sequences including flanks, the predictor was set to find a TMH of 10–30 residues within each sequence and give the ΔG_{app}; if no TMH was recognized, the sequence of lowest hydrophobicity was predicted using the "length correction" and "allow subsequence" options.

Evolutionary domain designations of TMHs were based on the cellular localization annotation found in the OPM database.[21] The 21 categories for cellular localization found were divided into three groups: Bacteria, Archaea, and Eukarya. Functional class designations were based on manual assessment of the annotation in PDBsum[22] and the original literature. Relative surface accessibility was calculated by NACCESS 2.1.139§[23] using the default probe size of 1.4 Å.[24] Statistical significance was tested with Welch two-sample *t* test and Fisher's exact test as implemented in R‖.[25]

The expected insertion efficiency, expressed as predicted change in free energy upon insertion, ΔG_{app}, was calculated for all transmembrane helices of the proteins in OPM, using the ΔG predictor¶. To allow for some flexibility in the structure, five flanking residues on each side were added to the structurally defined TMH. If any sequence of 19–23 residues in length within this segment had a lower predicted ΔG_{app} than the structural TMH, this other sequence was the one considered in the study. In our investigations, a sequence was considered being marginally hydrophobic if predicted to have $\Delta G_{app} > 1.4$ kcal/mol, corresponding to an apparent experimental insertion efficiency of less than 8%.

Acknowledgements

We thank Dr. B. Dobberstein, Heidelberg University, for providing rough microsomes. We are also grateful to Dr. D. Daley for the *E. coli* genes *acrB*, *amtB*, *btuC*, *clcA*, *cyoB*, *cyoC*, *emrD*, *glpT*, *lacY*, and *nhaA*; to Dr. H. Kim for the *S. cerevisiae rip1* gene; to Dr. J.K. Lanyi for the *H. salinarium bop* gene; to Dr. R. McKinnon for the *R. norvegicus* BK2; to Dr. T.A. Rapoport for the *M. jannaschii secY*; and to Dr. W. Skach for hsAQP1. This work was supported by grants from the Swedish Cancer Foundation (G.v. H., I.N.), the Swedish Foundation for Strategic Research (M.L.-B., A.K., S.C., A.E., I.N., G.v.H.), the Swedish Research Council (A.E., G.v.H., S.C.), Stiftelserna Riksförbundet Cystisk Fibros Forskningsfond and Anders Malmstens Minnesfond (I.N., K.E.), Magnus Bergvalls Stiftelse (I.N., S.C.), Henrik Granholms Stiftelse (I.N.), Carl Tryggers Stiftelse (I.N., S.C.), the European Union 6th and 7th Framework Programs (BioSapiens, FP6-LSHG-CT-2003-503265, EMBRACE FP6-LSHG-CT-2004-512092, EDICT FP7-HEALTH-F4-2007-201924), and the Academy of Finland (A.K.).

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.11.036

References

1. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M. *et al.* (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, **450**, 1026–1030.
2. Engelman, D. M., Chen, Y., Chin, C. N., Curran, A. R., Dixon, A. M., Dupuy, A. D. *et al.* (2003). Membrane protein folding: beyond the two stage model. *FEBS Lett.* **555**, 122–125.
3. Sadlish, H. & Skach, W. R. (2004). Biogenesis of CFTR and other polytopic membrane proteins: new roles for the ribosome–translocon complex. *J. Membr. Biol.* **202**, 115–126.
4. Pitonzo, D. & Skach, W. R. (2006). Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim. Biophys. Acta*, **1758**, 976–988.
5. Sato, Y., Sakaguchi, M., Goshima, S., Nakamura, T. & Uozumi, N. (2002). Integration of Shaker-type K^+ channel, KAT1, into the endoplasmic reticulum membrane: synergistic insertion of voltage-sensing segments, S3-S4, and independent insertion of pore-forming segments, S5-P-S6. *Proc. Natl Acad. Sci. USA*, **99**, 60–65.
6. Lundin, C., Kim, H., Nilsson, I., White, S. & von Heijne, G. (2008). Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N_{in}–C_{out} and N_{out}–C_{in} transmembrane helices. *Proc. Natl Acad. Sci. USA*, **105**, 15702–15707.
7. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H. *et al.* (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, **433**, 377–381.
8. Johansson, M., Nilsson, I. & von Heijne, G. (1993). Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. *Mol. Gen. Genet.* **239**, 251–256.
9. Kuroiwa, T., Sakaguchi, M., Omura, T. & Mihara, K. (1996). Reinitiation of protein translocation across the endoplasmic reticulum membrane for the topogenesis of multispanning membrane proteins. *J. Biol. Chem.* **271**, 6423–6428.
10. Heinrich, S., Mothes, W., Brunner, J. & Rapoport, T. (2000). The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell*, **102**, 233–244.
11. van den Berg, B., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature*, **427**, 36–44.

§ http://www.bioinf.manchester.ac.uk/naccess
‖ http://www.r-project.org/
¶ Available at: http://www.cbr.su.se/DGpred

12. Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. & Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. *Proc. Natl Acad. Sci. USA*, **105**, 7177–7181.
13. Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I. & von Heijne, G. (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA*, **105**, 4127–4132.
14. von Heijne, G. (1986). The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027.
15. Freites, J. A., Tobias, D. J., von Heijne, G. & White, S. H. (2005). Interface connections of a transmembrane voltage sensor. *Proc. Natl Acad. Sci. USA*, **102**, 15059–15064.
16. Johansson, A. C. & Lindahl, E. (2006). Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. *Biophys. J.* **91**, 4450–4463.
17. Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L. *et al.* (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J. Mol. Biol.* **387**, 1153–1164.
18. Buck, T. M., Wagner, J., Grund, S. & Skach, W. R. (2007). A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. *Nat. Struct. Mol. Biol.* **14**, 762–769.
19. Zimmer, J., Nam, Y. & Rapoport, T. A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature*, **455**, 936–943.
20. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.
21. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. (2006). OPM: orientations of proteins in membranes database. *Bioinformatics*, **22**, 623–625.
22. Laskowski, R. A., Chistyakov, V. V. & Thornton, J. M. (2005). PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.* **33**, D266–D268.
23. Hubbard, S. J., Campbell, S. F. & Thornton, J. M. (1991). Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J. Mol. Biol.* **220**, 507–530.
24. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
25. R Development Core Team. (2005). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria; http://www.R-project.org.