

Spiess, M., Junne, T., and Janoschke, M. (2019) Protein J. 38, 306-3016  
Final author manuscript

# Membrane protein integration and topogenesis at the ER

**Martin Spiess<sup>1</sup>, Tina Junne and Marco Janoschke**

Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

<sup>1</sup>Corresponding author: e-mail: martin.spiess@unibas.ch, +41-61-2072164

ORCID: 0000-0001-7139-0550

## Abstract

Most membrane proteins are composed of hydrophobic  $\alpha$ -helical transmembrane segments and are integrated into the lipid bilayer of the endoplasmic reticulum by the highly conserved Sec61 translocon. With respect to the integration mechanism, three types of transmembrane segments can be distinguished – the signal, the stop-transfer sequence, and the re-integration sequence – which in linear succession can account for all kinds of membrane protein topologies. The transmembrane orientation of the initial signal and to a weaker extent also of downstream transmembrane segments is affected by charged flanking residues according to the so-called positive-inside rule. The main driving force for transmembrane integration is hydrophobicity. Systematic analysis suggested thermodynamic equilibration of each peptide segment in the translocon with the membrane as the underlying mechanism. However, there is evidence that integration is not entirely sequence-autonomous, but depends also on the sequence context, from very closely spaced transmembrane segments to the folding state and properties of neighboring sequences. Topogenesis is even influenced by accessory proteins that appear to act as intramembrane chaperones.

**Keywords:** Sec61 / signal sequence / translocon / transmembrane helix

## Abbreviations

ER, endoplasmic reticulum; N<sub>cyt</sub>/C<sub>exo</sub>, cytoplasmic N- and exoplasmic C-terminus; N<sub>exo</sub>/C<sub>cyt</sub> exoplasmic N- and cytoplasmic C-terminus; TM, transmembrane; TRAM, translocating chain-associated membrane protein.

## Most membrane proteins are made of transmembrane helices

Approximately one third of our genes encodes membrane proteins. While lipid bilayers separate aqueous compartments – the inside and outside of cells and intracellular compartments –, membrane proteins connect them by mediating specific transport of molecules across that barrier. Other membrane proteins localize enzyme activities to the lipid surface, connect membranes to the cytoskeleton, or are involved in membrane traffic etc.

Structurally, the vast majority of integral membrane proteins are composed of *transmembrane (TM)  $\alpha$ -helices* embedded in the lipid bilayer, from a single membrane-spanning helix to several or even many helices bundled together. TM helices consist largely of hydrophobic amino acids with the side chains exposed to the hydrophobic core of the membrane and the hydrophilic peptide backbone hidden inside. The thickness of the fatty acyl chain region of the lipid bilayer of  $\sim 3$  nm defines the required length of a TM helix to be  $\sim 20$  residues ( $1.5$  Å progression per amino acid). With 3.6 amino acids per turn, such a helix consists of 5–6 turns. TM helices are frequently not perpendicular to the membrane, but tilted and thus somewhat longer.

As an example for helix-bundle membrane proteins, the Sec61/SecY translocon complex is shown in Figure 1A, i.e. the machinery that mediates integration of helical membrane proteins. It consists of three subunits conserved in all kingdoms: SecYEG in bacteria, SecYE $\beta$  in archaea (shown here), and Sec61 $\alpha/\gamma/\beta$ , respectively, in eukaryotes. The main  $\alpha$  subunit (SecY/61 $\alpha$ ) is a 10-helix protein bundled together with a single-helix  $\beta$  subunit (two TM helices in bacterial SecG) on one side and a  $\gamma$  subunit (SecE/61 $\gamma$ ) with a TM helix and an amphipathic helix on the membrane surface clasping two other sides. The surface embedded in the lipid is mainly apolar, in contrast to the parts exposed above or below the membrane (Fig. 1A, surface structure colored in green for hydrophilicity). The  $\alpha$  subunit produces a central pore for polypeptide translocation that is closed in the idle state by a constriction ring of 6 mainly hydrophobic residues and a luminal short plug helix. TM segments 2 and 7 form a lateral gate that provides a translocating chain access to the membrane interior.

To identify TM domains, hydrophobicity plots, where the average hydrophobicity of a window of 11–15 residues throughout a protein sequences is plotted [1], have proven very useful (Fig. 1B). Hydrophobic peptide stretches inside globular, soluble proteins are rarely long enough to be mistaken for TM segments. Since the hydrophobic effect is also the primary contribution to the internal packing of membrane proteins, even TM helices inside a bundle are mostly apolar all around.

The alternative structural principle for membrane proteins is the  *$\beta$ -barrel*, an anti-parallel closed  $\beta$ -sheet of 8–26  $\beta$ -strands of  $\sim 11$  residues each (6–22, depending on the tilt) [2, 3], forming a pore with more or less specificity (*porins*).  $\beta$ -Barrel membrane proteins are strictly limited to the outer membranes of bacteria, mitochondria, and chloroplasts. To be embedded in the bilayer, only every second amino acid side chain in a  $\beta$ -strand pointing to the outside of the barrel needs to be hydrophobic. As a result, their sequences are not sufficiently hydrophobic to integrate as TM helices. Indeed, bacterial porins are translocated by the SecYEG translocon as secretory proteins into the periplasm from where they are integrated into the outer membrane by insertases that are themselves  $\beta$ -barrels. It has been proposed that successive  $\beta$ -strands of nascent  $\beta$ -barrel proteins intercalate between the first and last strands of the insertase until they bud out as independent  $\beta$ -barrels [4, 5]. By analogy, TM helices of nascent helical membrane proteins intercalate into the lateral gate of the SecY/61 translocon to be successively released to form an independent helix bundle. Indeed, recent structures of translocons containing signal or TM domains illustrate exactly this intercalation between the helices of the lateral gate [6–8].

## Three distinct processes of TM segment integration

The first sequence of a membrane protein to engage with the translocon is the *signal* sequence. After recruitment of signal recognition particle (SRP), targeting of the SRP–nascent chain–ribosome complex to the SRP receptor at the endoplasmic reticulum (ER) membrane, and transfer of the complex to the translocon, the signal somehow interacts with the Sec61 translocon by inserting itself into the translocation pore. This insertion process is what several translocation inhibitors, such as cotransin, decatransin, and apratoxin, have been shown to block [9–11].

These inhibitors either block access of the signal to the translocon and/or they stabilize and lock the closed state of Sec61.

Three classes of signal sequences can be distinguished, as illustrated in Fig.2 (①a–c). Classical *cleavable signals* (①a), initially discovered by Blobel and Dobberstein [12], insert into the translocon as a loop to place the downstream sequence into the translocation pore. This was illustrated by a cryo-electron microscopy structure of ribosome-bound translocon containing a stalled nascent chain [6]. The signal's C-terminal end is exposed to ER lumen and cleaved by signal peptidase, generating a new N-terminus on the exoplasmic side. *Signal-anchors* insert in the same way in an  $N_{\text{cyt}}/C_{\text{exo}}$  orientation (cytosolic N- and exoplasmic C-terminus), but are not cleaved and thus remain in the bilayer as a TM anchor (①b). They generally have a longer hydrophobic core than cleavable signals to comfortably span the membrane and they are not necessarily positioned at the very N-terminus, but may be preceded by any length of polypeptide. *Reverse signal-anchors* insert in the opposite orientation and initiate translocation of their N-terminal end, finally anchoring the protein in the opposite  $N_{\text{exo}}/C_{\text{cyt}}$  orientation (①c).

As a polypeptide is passing through the translocon, a hydrophobic TM segment will stop further translocation as a so-called *stop-transfer* sequence by exiting the pore laterally into the lipid bilayer (Fig. 2, ②). The downstream sequence will be directly synthesized through a gap between the ribosome and the translocon into the cytosol. A subsequent TM segment, a *re-integration* sequence (③), again integrates itself into the membrane via the translocon in an  $N_{\text{cyt}}/C_{\text{exo}}$  orientation, inserting its immediate downstream sequence into the pore for translocation. This process is similar to initial insertion by a signal-anchor, except that the translocon might already be in an open state and possibly still be associated with the preceding stop-transfer sequence.

Three types of single-spanning membrane proteins may thus be produced by a cleavable signal and a stop-transfer sequence ( $N_{\text{exo}}/C_{\text{cyt}}$ ; type I or type Ia), by a signal-anchor sequence ( $N_{\text{cyt}}/C_{\text{exo}}$ ; type II), or by a reverse signal-anchor ( $N_{\text{exo}}/C_{\text{cyt}}$ ; type III or type Ia). For multi-spanning (polytopic) membrane proteins, it has been proposed already in 1980 by Blobel that they achieve their final topology by an additional succession of alternating stop-transfer and re-integration sequences (or, as he called them, "internal signal sequences") [13].

### The positive-inside rule and N-terminal folding determine signal orientation

The signal sequence engages with the translocon to insert one of its flanking sequences into the translocation pore for transfer across the membrane. While cleavable signals and signal-anchors translocate their C-terminal end, the N-terminal sequence is translocated by reverse signal-anchors. Cleavable signals typically carry positive charges in the n-region that remains cytoplasmic [14]. Arginines and lysines were found to be statistically enriched in the cytoplasmic portions of membrane proteins as a general phenomenon, the *positive-inside rule*, and serves as a useful criterion in topology prediction [15, 16]. In particular, it also holds true for the sequences flanking signal-anchors and reverse signal-anchors [17]. Mutation of flanking charges resulted in protein inversion in both directions (e.g. [18-20]) demonstrating a causal relationship between flanking charges and signal orientation. However, charge inversion was not generally sufficient to produce an uniform topology; additional factors thus also influence signal orientation at the translocon.

The available structures of the Sec61/Y translocon confirm that the pore is too narrow to allow folded domains beyond individual helices to pass through. In cotranslational translocation, the nascent polypeptide is largely kept in an unfolded state as it passes through the ribosome exit tunnel to the translocation pore. This is not the case for the N-terminal sequence preceding the signal sequence, which emerges from the ribosome into the cytosol before SRP-dependent membrane targeting. Rapid folding of N-terminal domains inhibit their translocation, overriding the positive-inside rule [21]. Most reverse signal-anchor proteins have rather short N-terminal domains, for example the synaptotagmins with up to ~60 residues, but there are exceptions of  $N_{\text{exo}}/C_{\text{cyt}}$  single-spanning proteins with N-domains of more than 100 residues (e.g. 138 residues for the ectodysplasin A2 receptor [TNR27\_HUMAN in uniprot.org] and 242 residues for pro-neuregulin-1 [NRG-1\_HUMAN]). Apparently, this sequences do not sufficiently fold in the cytoplasm to hinder their translocation, probably also because its disulfide bonds are only formed in the ER lumen.

## The mechanism of signal orientation

The origin of the positive-inside rule in eukaryotes is not clear. There is no general membrane potential across the ER membrane that might contribute. An attractive mechanism is the cytosolic retention of the more positive end of the signal by negative charges at or near the translocon, for example also by the net negative lipid headgroups. An alternative mechanism was discovered testing artificial signal-anchors composed of increasingly hydrophobic oligo-leucine stretches: the more hydrophobic the core of the signal, the more N-translocation is favored, for N-terminal signals even overriding the charge rule [22, 23]. Testing different amino acids either as homo-oligomers or as guest residues in an oligo-leucine host signal confirmed that it is side chain hydrophobicity that promotes N-translocation [24].

How this might happen was suggested by the observation that the final topologies of model proteins with very hydrophobic N-terminal signals depended on the length of the protein. An N-terminal signal-anchor with a generic of 22-leucine h-domain inserted to a large fraction with an  $N_{\text{exo}}/C_{\text{cyt}}$  orientation, despite a positive N-terminus. This fraction (i.e. N-translocation) was highest for a short protein of ~100 residues following the signal sequence and decreased up to ~300 residues [25]. This result indicated that N-terminal signals initially insert with the N-terminus in the ER lumen and then invert orientation until protein synthesis is terminated or until the reorientation process ends, possibly due to lipid integration of the TM signal (Fig. 3A). Consistent with this interpretation, the fraction of C-translocated signal-anchors increased for each construct, when more time until protein completion was provided by slowing down translation rate with cycloheximide. Signal inversion was shown to be driven by N-terminal positive charge according to the positive-inside rule and inhibited by hydrophobicity (likely by stabilizing the initial orientation bound to translocon and lipid). Most or all natural cleavable signals and signal-anchors are less hydrophobic and thus invert within seconds, long before translation is completed. While this model was originally derived from endpoint topology analysis of model proteins expressed *in vivo* in COS-1 cells [25], the process of head-on insertion and inversion of a natural signal-anchor was corroborated more directly by *in vitro* translation/ translocation into dog pancreas microsomes using arrested nascent chains and biochemical analysis [26].

Not surprisingly, initial head-on insertion is only possible for N-terminal signals: with n-domains longer than 20 residues time/length-dependent reorientation of signals was not observed anymore [27]. Internal signals therefore position themselves according their flanking charges, before they insert into the translocon and open it (Fig. 3B). The apparent difference in the integration mechanism of N-terminal and internal signals may be the cause of the observed difference in sensitivity to the translocon inhibitor mycolactone, which blocks  $N_{\text{cyt}}/C_{\text{exo}}$  integration and inversion, but not  $N_{\text{exo}}/C_{\text{cyt}}$  integration [28].

What is the contribution of the Sec61 translocon in orienting signal sequences according to the positive-inside rule? Mutation of candidate charges in yeast Sec61p identified three residues (R67 and R74 in the plug on the luminal side, and E382 on the cytoplasmic side) that, when mutated, weaken C-translocation of a diagnostic N-terminally positive signal and N-translocation of a C-terminally positive one [29]. A screen for mutations in Sec61p that affect signal orientation, however, revealed a complex situation [30]. Three classes of mutations with distinct effects on different substrates could be distinguished, one of them with the same phenotype of less efficiently retaining the positively charged flanking region in the cytosol. These mutations localize to different positions in the Sec61 protein and do not only involve charged residues. Almost all of them have a *prl* (protein localization) phenotype initially described in bacteria as suppressors of signal sequence mutations [31]. They appear to destabilize the closed state of the translocon causing it to open more easily and also with very weak signals [32-34]. Rapid or "premature" pore opening, before the signal had time to properly position itself may thus seem to weaken the positive-inside rule [30]. However, a direct role for the positively charged residues on the plug in signal orientation cannot be excluded.

### Topogenic information in multi-spanning proteins

Once the signal sequence has adopted its orientation, that of the downstream TM segments is defined only by their relative positions in the protein. The activity of stop-transfer and re-integration sequences appears not to be very specific. For example, at least five of the seven TM segments of rhodopsin (a reverse signal-anchor followed by three re-integration and three stop-transfer sequences) were able to function as signal sequences [35]. Conversely, chimeric proteins constructed from two to four identical copies of a signal-anchor integrated readily as two- to four-fold membrane-spanning proteins, respectively [36, 37]: the even-numbered signal-anchors inserted as stop-transfer sequences in an inverted orientation, against the charge rule.

While this is true for TM segments generously separated from each other in model proteins, it is not necessarily the case in natural multi-spanning membrane proteins, indicating that there are additional factors defining topology. One might have expected that mutating a signal-anchor to a reverse signal-anchor would result in inversion of the entire protein. This was not observed, when it was experimentally tested with the glucose transporter GLUT1. Instead, the second TM segment did not integrate in a membrane-spanning manner, leaving the downstream topology unchanged [38]. Similarly, mutation of positive charge clusters in short cytoplasmic loops caused both neighboring TM segments to be translocated, while upstream and downstream of this disturbance, the topology remained as in the wild-type [39]. Such topological "frustration" was also observed in artificial multi-spanning proteins with mismatched charge distribution in *E.coli* [40].

Of particular importance is here that TM segments in multi-spanning proteins are often very closely spaced. Competition between two conflicting topogenic elements, a cleaved signal and an internalized signal-anchor, was observed with intervening sequences  $\leq 60$  residues [41]. The shorter the connecting peptide, the more extensive the topology was rearranged. In natural proteins, TM segments are frequently separated by only a few amino acid, forming the turn of a helical hairpin. In such a situation, the first TM domain cannot possibly (re-)integrate into the membrane independently of the second. They rather act together in a distinct hairpin insertion process (illustrated in Fig. 2, bottom). Indeed, it was shown that helical transmembrane hairpins can fold already in the exit vestibule of ribosome [42]. In addition, pre-assembly of TM helices containing residues of opposite charge may reduce the energetic cost of their integration.

### Membrane integration of stop-transfer and re-integration sequences

While it is self-evident that TM helices must be hydrophobic for membrane integration, a systematic analysis of the contribution of every amino acid to the process (the "molecular code for TM helix recognition by the Sec61 translocon") was performed by von Heijne and coworkers for stop-transfer integration [43, 44]. Mildly hydrophobic so-called H-segments were created based on a 19-residue oligo-alanine sequence (Fig. 4A). By exchanging increasing numbers of alanines to other amino acids, a large number of H-segments were produced and tested in a reporter protein for their ability to integrate into the bilayer and to stop polypeptide transfer (Fig. 4B–D). The results suggested membrane insertion to be a thermodynamic equilibration process between the lipid and the translocation pore. This allows to calculate the apparent free energy contribution ( $\Delta G_{app}$ ; Fig. 4E) of each amino acid and at any position in an H-segment to membrane insertion and thus prediction of transmembrane segments, at least for single TM domains, where the process is not complicated by specific interactions with neighboring TM helices (as analyzed in ref. [45]).

The result was a "biological hydrophobicity scale" of amino acids [43] that largely parallels biophysical scales, except that it appears compressed and shifted [46–48]. In part, this may be because equilibration occurs between lipid and translocon, and not between lipid and free solution (Fig. 4F). The translocon interior is a narrow space of low hydration, mainly because of its constriction ring of six mostly apolar residues forming a gasket. Mutation of these residues to more polar ones indeed enhanced integration of H-segments [49, 50], perfectly in line with equilibration. By the residues lining the inside of the pore, the translocon thus defines the hydrophobicity threshold for membrane integration.

In the simplest model, the nascent chain, as it moves through the pore at the speed of translation, has access to the lipid through the lateral gate. At an elongation rate of  $\sim 5$  residues per second, each segment is in register

with the membrane in the order of  $\sim 1$  s, during which equilibrium is reached. The probability of the two states directly reflects that of integration and translocation (Fig. 4F). Yet, an H-segment – particularly one of intermediate hydrophobicity – will find an energetic minimum in the translocon's gate between the lipid and the partially water-filled pore, and also vertically between the aqueous environments of ER lumen and cytosol (Fig. 4G). Indeed, TM segments were found positioned in the gate in molecular dynamics simulations and cryo-electron microscopy structures (e.g. refs. [7, 50]). Using constructs with a SecM stalling sequence to measure the force acting on the nascent chain, two force peaks were observed when the TM segment reached into the pore and then into the lipid [51]. Integration vs. translocation is determined by the relative rates of exit from gate position either into the lipid or the lumen. With increasing length of the polypeptide chain accumulating between ribosome and translocon, the reversibility of the exit is rapidly decreasing (Fig. 4G). The membrane acts as an entropic trap for the TM segment and chaperone binding to the lumenally exposed sequence as a ratchet. Consistent with this view, the conformational properties of a sequence up to 100 residues downstream of the H-segment affected the hydrophobicity threshold of integration [52]. Integration was facilitated, when the sequence was flexible and extended, and inhibited when compact, reflecting the gain in entropy permitted by the downstream sequence. This result indicates that sequences do not define autonomously their integration behavior, but are influenced by their sequence context. Similarly, one might speculate that the folding and chaperone binding properties of the upstream sequence might influence the rate of exit into the ER lumen. This context dependence of membrane integration is at least in part the explanation, why the hydrophobicity thresholds determined with identical leucine-containing H-segments in different model proteins and expression systems produced different results (Fig. 5).

While the process of stop-transfer integration has been extensively studied, the principles of re-integration remain to be dissected. As reported by Lundin et al. [53], the molecular code for re-integration generally parallels that for stop-transfer integration, except that the overall hydrophobicity threshold is significantly lower. In a SecM pulling force experiment, Cymer et al. [54] found a re-integration TM segment to exert a weaker pulling force on an arrested chain compared to a stop-transfer segment, confirming that they engage with the translocon differently.

### Accessory factors regulating membrane protein biogenesis

The Sec61 translocon can associate with a complex of Sec62 and Sec63 – membrane proteins with two and three TM segments, respectively – and peripherally attached Sec71 and Sec72 [55, 56]. The luminal J-domain of Sec63 recruits the Hsc70 chaperone BiP (immunoglobulin binding protein) to capture the translocating chain and thus to provide vectoriality for post-translational translocation. Recent cryo-electron microscopy structures of the Sec61 translocon with the Sec63 complex demonstrate how the Sec61 channel is activated for post-translational protein translocation [57, 58]. However, Sec63 and BiP (Kar2p in yeast) were also found to be important for SRP-dependent, co-translational translocation [59-61]. Interestingly, the new structures show Sec63 complex bound to Sec61 in a manner blocking simultaneous binding of a ribosome. How to reconcile this with the effects of the Sec62/63 complex in cotranslational translocation is not clear at present.

The Sec61 translocon is furthermore accompanied by several auxiliary proteins that assist integration of subsets of membrane proteins by mostly unclear mechanisms. Translocating chain-associated membrane protein (TRAM) was initially shown to be required for integration of signals with short n- and h-domains in a reconstituted *in vitro* system [62]. Recently, TRAM2 was discovered to ensure correct insertion of the reverse signal-anchor of a four-TM protein TM4SF20, an inhibitor of cleavage of the membrane bound transcription factor CREB3L1 [63]. TRAM proteins contain a potential ceramide-binding domain. In the presence of ceramide or in the absence of TRAM2, the reverse signal-anchor of TM4SF20 inserts as an N<sub>cyt</sub>/C<sub>exo</sub> signal-anchor, in this case even resulting in inversion of the topology of all three downstream TM domains.

Recently, the ER membrane protein complex EMC was identified to contribute to membrane protein folding [64]. It was specifically shown to be required for insertion of the first TM domain (the reverse signal-anchor) of some G-protein coupled receptors [65]. In a systematic proteomic approach, the EMC was found to be involved

in the topogenesis of multi-spanning proteins in general, acting as a membrane chaperone especially for mildly hydrophobic TM domains [66].

The contribution of different membrane lipids on protein integration is poorly studied in eukaryotes, but there is evidence in bacteria demonstrating surprisingly strong effects on topology particularly the lactose permease LacY (see accompanying review on "The role of lipids in membrane protein biogenesis" by W. Dowhan). Lipid composition is certainly very different in some of the system studied in Figure 5 and might contribute to the observed differences. Furthermore, it seems likely that additional factors –substrate specific or even general – will be discovered to contribute to protein integration into the membrane, a process that goes beyond partitioning of hydrophobic surfaces into the bilayer.

### **Acknowledgments**

Our work was supported by Grant 31003A-182519 from the Swiss National Science Foundation.

## References

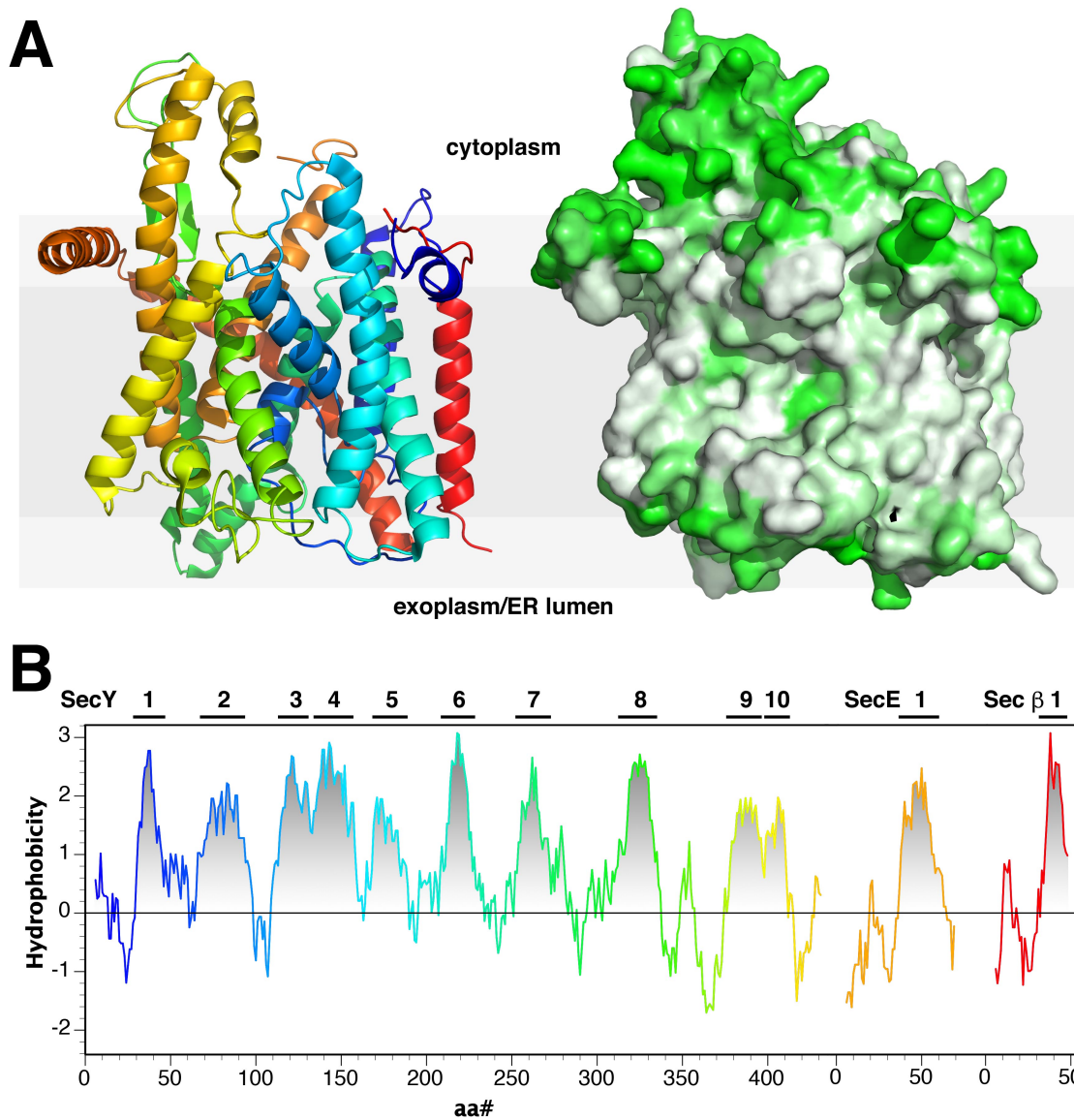
1. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132. doi: 10.1016/0022-2836(82)90515-0
2. Wimley WC (2002) Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein Sci* 11:301–312. doi: 10.1110/ps.29402
3. Schiffrin B, Brockwell DJ, Radford SE (2017) Outer membrane protein folding from an energy landscape perspective. *BMC Biol* 15:123. doi: 10.1186/s12915-017-0464-5
4. Gruss F, Zähringer F, Jakob RP, et al (2013) The structural basis of autotransporter translocation by TamA. *Nat Struct Mol Biol* 20:1318–1320. doi: 10.1038/nsmb.2689
5. Höhr AIC, Lindau C, Wirth C, et al (2018) Membrane protein insertion through a mitochondrial  $\beta$ -barrel gate. *Science* 359:eaah6834. doi: 10.1126/science.aah6834
6. Park E, Ménétret J-F, Gumbart JC, et al (2014) Structure of the SecY channel during initiation of protein translocation. *Nature* 506:102–106. doi: 10.1038/nature12720
7. Gogala M, Becker T, Beatrix B, et al (2014) Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature* 506:107–110. doi: 10.1038/nature12950
8. Voorhees RM, Hegde RS (2016) Structure of the Sec61 channel opened by a signal sequence. *Science* 351:88–91. doi: 10.1126/science.aad4992
9. MacKinnon AL, Paavilainen VO, Sharma A, et al (2014) An allosteric Sec61 inhibitor traps nascent transmembrane helices at the lateral gate. *Elife* 3:e01483. doi: 10.7554/eLife.01483
10. Junne T, Wong J, Studer C, et al (2015) Decatransin, a novel natural product inhibiting protein translocation at the Sec61/SecY translocon. *J Cell Sci* 128:1217–1229. doi: 10.1242/jcs.165746
11. Paatero AO, Kellosalo J, Dunyak BM, et al (2016) Apratoxin Kills Cells by Direct Blockade of the Sec61 Protein Translocation Channel. *Cell Chem Biol* 23:561–566. doi: 10.1016/j.chembiol.2016.04.008
12. Blobel G, Dobberstein B (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol* 67:835–851. doi: 10.1083/jcb.67.3.835
13. Blobel G (1980) Intracellular protein topogenesis. *Proc Natl Acad Sci USA* 77:1496–1500.
14. Heijne von G (1990) The signal peptide. *J Membr Biol* 115:195–201.
15. 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.
16. Nilsson J, Persson B, Heijne von G (2005) Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* 60:606–616. doi: 10.1002/prot.20583
17. Hartmann E, Rapoport TA, Lodish HF (1989) Predicting the orientation of eukaryotic membrane-spanning proteins. *86:5786–5790*.
18. Szczesna-Skorupa E, Kemper B (1989) NH<sub>2</sub>-terminal substitutions of basic amino acids induce translocation across the microsomal membrane and glycosylation of rabbit cytochrome P450IIC<sub>2</sub>. *J Cell Biol* 108:1237–1243.
19. Beltzer JP, Fiedler K, Fuhrer C, et al (1991) Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. *J Biol Chem* 266:973–978.



20. Parks GD, Lamb RA (1991) Topology of eukaryotic type II membrane proteins: importance of N-terminal positively charged residues flanking the hydrophobic domain. *Cell* 64:777–787. doi: 10.1016/0092-8674(91)90507-U
21. Denzer AJ, Nabholz CE, Spiess M (1995) Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO J* 14:6311–6317.
22. Sakaguchi M, Tomiyoshi R, Kuroiwa T, et al (1992) Functions of signal and signal-anchor sequences are determined by the balance between the hydrophobic segment and the N-terminal charge. *Proc Natl Acad Sci USA* 89:16–19. doi: 10.1073/pnas.89.1.16
23. Wahlberg JM, Spiess M (1997) Multiple determinants direct the orientation of signal-anchor proteins: the topogenic role of the hydrophobic signal domain. 137:555–562.
24. Rösch K, Naeher D, Laird V, et al (2000) The topogenic contribution of uncharged amino acids on signal sequence orientation in the endoplasmic reticulum. *J Biol Chem* 275:14916–14922. doi: 10.1074/jbc.M000456200
25. Goder V, Spiess M (2003) Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* 22:3645–3653. doi: 10.1093/emboj/cdg361
26. Devaraneni PK, Conti B, Matsumura Y, et al (2011) Stepwise Insertion and Inversion of a Type II Signal Anchor Sequence in the Ribosome-Sec61 Translocon Complex. 146:134–147. doi: 10.1016/j.cell.2011.06.004
27. Kocik L, Junne T, Spiess M (2012) Orientation of internal signal-anchor sequences at the Sec61 translocon. *J Mol Biol* 424:368–378. doi: 10.1016/j.jmb.2012.10.010
28. McKenna M, Simmonds RE, High S (2017) Mycolactone reveals the substrate-driven complexity of Sec61-dependent transmembrane protein biogenesis. *J Cell Sci* 130:1307–1320. doi: 10.1242/jcs.198655
29. Goder V, Junne T, Spiess M (2004) Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol Biol Cell* 15:1470–1478. doi: 10.1091/mbc.E03-08-0599
30. Junne T, Schwede T, Goder V, Spiess M (2007) Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. *J Biol Chem* 282:33201–33209. doi: 10.1074/jbc.M707219200
31. Emr SD, Hanley-Way S, Silhavy TJ (1981) Suppressor mutations that restore export of a protein with a defective signal sequence. 23:79–88.
32. Tam PCK, Maillard AP, Chan KKY, Duong F (2005) Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J* 24:3380–3388. doi: 10.1038/sj.emboj.7600804
33. Saparov SM, Erlandson K, Cannon K, et al (2007) Determining the Conductance of the SecY Protein Translocation Channel for Small Molecules. *Mol Cell* 26:501–509. doi: 10.1016/j.molcel.2007.03.022
34. Flower (2007) The SecY translocation complex: convergence of genetics and structure. *Trends Microbiol.* doi: 10.1016/j.tim.2007.03.001
35. Audigier Y, Friedlander M, Blobel G (1987) Multiple topogenic sequences in bovine opsin. *Proc Natl Acad Sci USA* 84:5783–5787. doi: 10.1073/pnas.84.16.5783
36. Wessels HP, Spiess M (1988) Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. 55:61–70.
37. Lipp J, Flint N, Haeuptle MT, Dobberstein B (1989) Structural requirements for membrane assembly of proteins spanning the membrane several times. *J Cell Biol* 109:2013–2022. doi: 10.1083/jcb.109.5.2013
38. Sato M, Hresko R, Mueckler M (1998) Testing the charge difference hypothesis for the assembly of a eucaryotic multispanning membrane protein. *J Biol Chem* 273:25203–25208. doi: 10.1074/jbc.273.39.25203

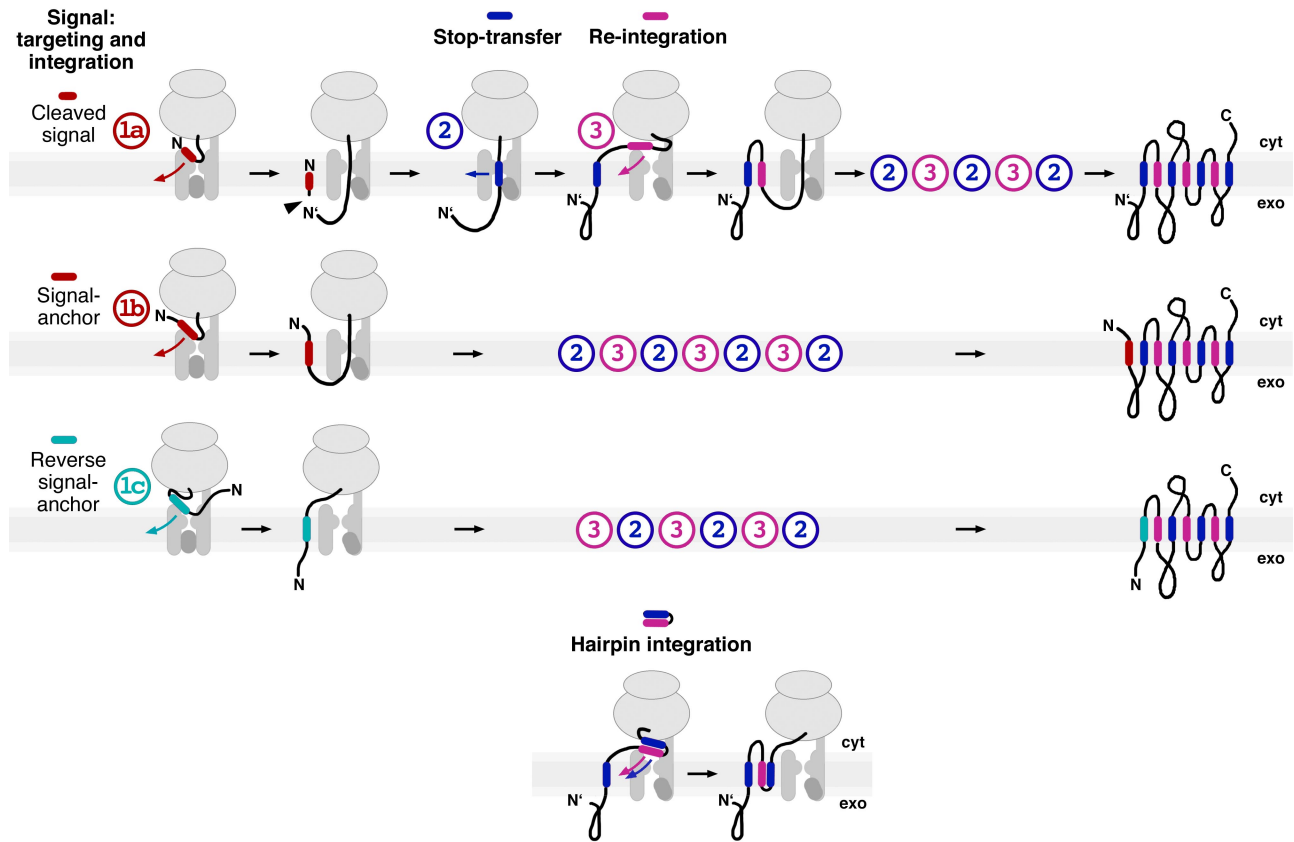
39. Sato M, Mueckler M (1999) A conserved amino acid motif (R-X-G-R-R) in the Glut1 glucose transporter is an important determinant of membrane topology. *J Biol Chem* 274:24721–24725. doi: 10.1074/jbc.274.35.24721
40. Gafvelin G, Heijne von G (1994) Topological “frustration” in multispanning E. coli inner membrane proteins. *Cell* 77:401–412. doi: 10.1016/0092-8674(94)90155-4
41. Goder V, Bieri C, Spiess M (1999) Glycosylation can influence topogenesis of membrane proteins and reveals dynamic reorientation of nascent polypeptides within the translocon. 147:257–266.
42. Tu L, Khanna P, Deutsch C (2014) Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *J Mol Biol* 426:185–198. doi: 10.1016/j.jmb.2013.09.013
43. Hessa T, Kim H, Bihlmaier K, et al (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433:377–381. doi: 10.1038/nature03216
44. Hessa T, Meindl-Beinker NM, Bernsel A, et al (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* 450:1026–1030. doi: 10.1038/nature06387
45. Hedin LE, Ojemalm K, Bernsel A, et al (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* 396:221–229. doi: 10.1016/j.jmb.2009.11.036
46. MacCallum JL, Tieleman DP (2011) Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends Biochem Sci* 36:653–662. doi: 10.1016/j.tibs.2011.08.003
47. Schow EV, Freites JA, Cheng P, et al (2011) Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J Membr Biol* 239:35–48. doi: 10.1007/s00232-010-9330-x
48. Gumbart JC, Teo I, Roux B, Schulten K (2013) Reconciling the roles of kinetic and thermodynamic factors in membrane-protein insertion. *J Am Chem Soc* 135:2291–2297. doi: 10.1021/ja310777k
49. Junne T, Kocik L, Spiess M (2010) The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. *Mol Biol Cell* 21:1662–1670. doi: 10.1091/mbc.E10-01-0060
50. Demirci E, Junne T, Baday S, et al (2013) Functional asymmetry within the Sec61p translocon. *Proc Natl Acad Sci USA* 110:18856–18861. doi: 10.1073/pnas.1318432110
51. Ismail N, Hedman R, Schiller N, Heijne von G (2012) A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. 19:1018–1022. doi: 10.1038/nsmb.2376
52. Junne T, Spiess M (2017) Integration of transmembrane domains is regulated by their downstream sequences. *J Cell Sci* 130:372–381. doi: 10.1242/jcs.194472
53. Lundin C, Kim H, Nilsson I, et al (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. 105:15702–15707. doi: 10.1073/pnas.0804842105
54. Cymer F, Ismail N, Heijne von G (2014) Weak pulling forces exerted on Nin-orientated transmembrane segments during co-translational insertion into the inner membrane of Escherichia coli. *FEBS Lett* 588:1930–1934. doi: 10.1016/j.febslet.2014.03.050
55. Deshaies RJ, Sanders SL, Feldheim DA, Schekman R (1991) Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* 349:806–808. doi: 10.1038/349806a0
56. Panzner S, Dreier L, Hartmann E, et al (1995) Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. 81:561–570.

57. Itskanov S, Park E (2019) Structure of the posttranslational Sec protein-translocation channel complex from yeast. *Science* 363:84–87. doi: 10.1126/science.aav6740
58. Wu X, Cabanos C, Rapoport TA (2019) Structure of the post-translational protein translocation machinery of the ER membrane. *Nature* 566:136–139. doi: 10.1038/s41586-018-0856-x
59. Brodsky JL, Goeckeler J, Schekman R (1995) BiP and Sec63p are required for both co- and posttranslational protein translocation into the yeast endoplasmic reticulum. *92*:9643–9646.
60. Young BP, Craven RA, Reid PJ, et al (2001) Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo. *EMBO J* 20:262–271. doi: 10.1093/emboj/20.1.262
61. Lang S, Benedix J, Fedeles SV, et al (2012) Different effects of Sec61 $\alpha$ , Sec62 and Sec63 depletion on transport of polypeptides into the endoplasmic reticulum of mammalian cells. *J Cell Sci* 125:1958–1969. doi: 10.1242/jcs.096727
62. Voigt S, Jungnickel B, Hartmann E, Rapoport TA (1996) Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. *134*:25–35.
63. Chen Q, Denard B, Lee C-E, et al (2016) Inverting the Topology of a Transmembrane Protein by Regulating the Translocation of the First Transmembrane Helix. *Mol Cell* 63:567–578. doi: 10.1016/j.molcel.2016.06.032
64. Jonikas MC, Collins SR, Denic V, et al (2009) Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* 323:1693–1697. doi: 10.1126/science.1167983
65. Chitwood PJ, Juszkievicz S, Guna A, et al (2018) EMC Is Required to Initiate Accurate Membrane Protein Topogenesis. *Cell* 175:1507–1519.e16. doi: 10.1016/j.cell.2018.10.009
66. Shurtleff MJ, Itzhak DN, Hussmann JA, et al (2018) The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *Elife* 7:382. doi: 10.7554/eLife.37018
67. van den Berg B, Clemons WM, Collinson I, et al (2004) X-ray structure of a protein-conducting channel. *Nature* 427:36–44. doi: 10.1038/nature02218
68. Sommer N, Junne T, Kalies K-U, et al (2013) TRAP assists membrane protein topogenesis at the mammalian ER membrane. *Biochim Biophys Acta* 1833:3104–3111. doi: 10.1016/j.bbamcr.2013.08.018
69. Hessa T, Reithinger JH, Heijne von G, Kim H (2009) Analysis of transmembrane helix integration in the endoplasmic reticulum in *S. cerevisiae*. *J Mol Biol* 386:1222–1228.
70. Xie K, Hessa T, Seppälä S, et al (2007) Features of transmembrane segments that promote the lateral release from the translocase into the lipid phase. *Biochemistry* 46:15153–15161. doi: 10.1021/bi701398y
71. Ojemalm K, Botelho SC, Stüdle C, Heijne von G (2013) Quantitative analysis of SecYEG-mediated insertion of transmembrane  $\alpha$ -helices into the bacterial inner membrane. *J Mol Biol* 425:2813–2822. doi: 10.1016/j.jmb.2013.04.025



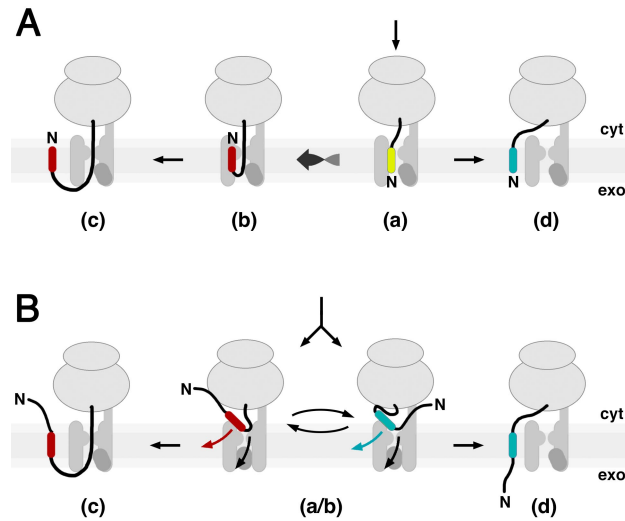
**Figure 1. Two basic types of membrane proteins:  $\alpha$ -helix bundles and  $\beta$ -barrels.**

A: As an example of a helix bundle membrane protein, the structure of the archaeal SecYE $\beta$  translocon from *Methanococcus janaschii* (1RHZ [67]) is shown as backbone ribbon (left) and as a surface representation colored green for hydrophilic residues. B: The corresponding hydrophobicity plot is shown according to Kyte and Doolittle [1] (11-residue window).



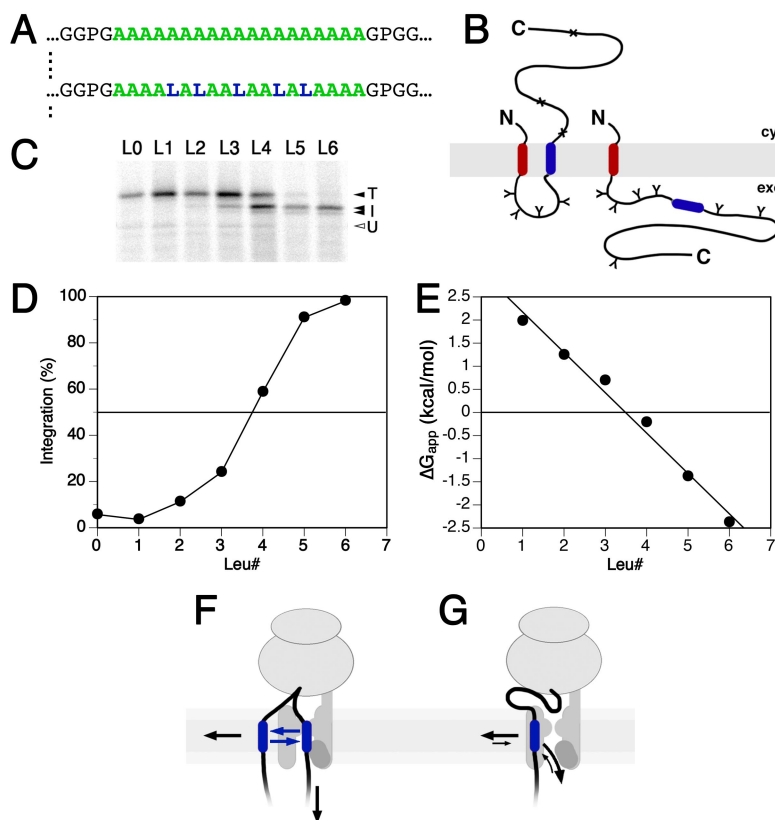
**Figure 2. Three types of TM integration: signal, stop-transfer, and re-integration.**

Topogenic sequences integrating in an N<sub>cyt</sub>/C<sub>exo</sub> orientation are shown in red for cleaved signal and signal-anchor, and pink for re-integration. TM segments integrating in an N<sub>exo</sub>/C<sub>cyt</sub> orientation are shown in turquoise for reverse signal-anchor and blue for stop-transfer. Successive stop-transfer and re-integration TM segments result in multi-spanning proteins, illustrated here for 7–8 TM proteins. For example, approximately two thirds of the seven-TM G-protein coupled receptors utilize a cleavable signal (①a②③②③②③②), the rest a reverse signal-anchor (①c③②③②③②③) to initiate topogenesis, as shown schematically. Below, hairpin integration is shown as an example that two closely spaced TM segments cannot integrate independently of each other.



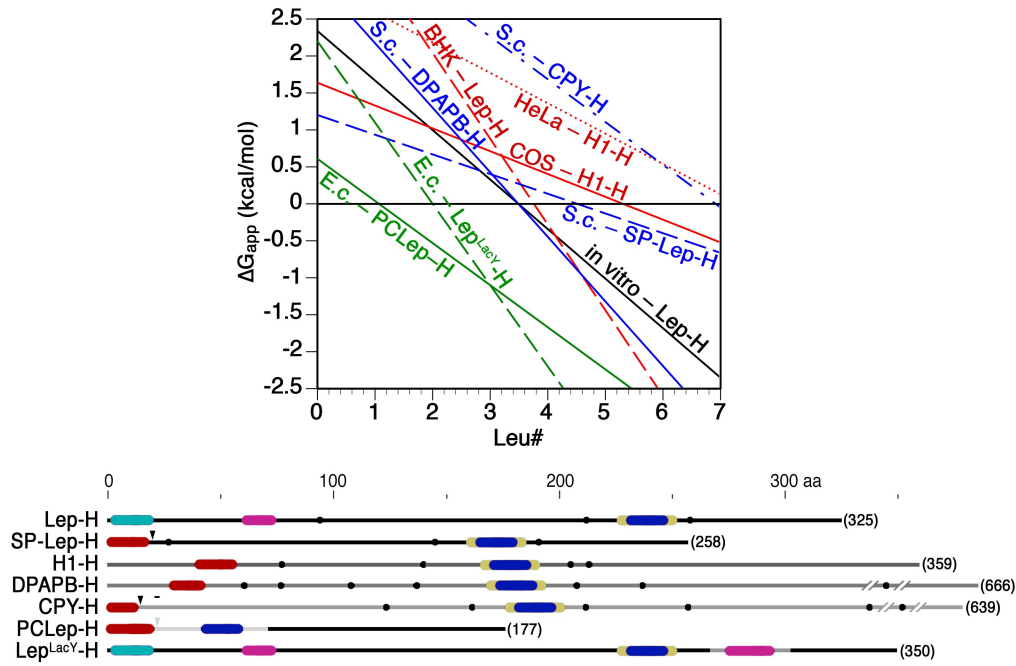
**Figure 3. Mechanistic model of signal orientation.**

A: N-terminal signals initially insert head-on in an  $N_{\text{exo}}/C_{\text{cyt}}$  orientation (a). Following the positive-inside rule, they either invert orientation to  $N_{\text{cyt}}/C_{\text{exo}}$  (b) to integrate as signal-anchors (c), or they retain the original direction and integrate as reverse signal-anchors (d). Inversion is slowed down by high signal hydrophobicity. B: Internal signals are hindered by their N-terminal sequence from head-on insertion and position themselves according to the positive-inside rule before pore opening and translocation. Destabilization of the translocon by *prl* mutations leads to integration before correct alignment and thus appears to weaken the charge rule.



**Figure 4. Membrane integration as thermodynamic equilibration.**

A: H-Segments are based on a 19-alanine guest sequence in which one or more alanines are replaced by other amino acids. In a simple hydrophobicity series, 0–6 leucine replacements were made (L0 and L5 are shown). B: Inserted in a model protein (here dipeptidylaminopeptidase B [50], H-segments (in blue) will integrate into the membrane or be translocated with different glycosylation (Y). C: Upon expression (here in yeast with 5-min [ $^{35}$ S]methionine labeling), the glycosylation pattern of the products after SDS-gel electrophoresis and autoradiography reveals the extent of H-segment translocation (T, full glycosylation) and integration (I, partial glycosylation; U, unglycosylated). D: Quantitation shows the hydrophobicity threshold for 50% integration to be slightly less than 4 leucines. E: Considering the results to be apparent equilibration constants  $K_{app} = I/T$ , apparent free energies for integration  $\Delta G_{app} = -RT \ln K_{app}$  can be calculated. F and G: Two model representations of the partitioning process of the H-segment. (Panels B–D were adapted from ref. [50]).



**Figure 5. Dependence of  $\Delta G_{app}$  of H-segment integration on the system and model protein.**

$\Delta G_{app}$  for stop-transfer integration is plotted against the number of leucines in the H-segments (as in Fig. 4A) analyzed in different expression systems and model proteins: in vitro - Lep-H [43]; BHK - Lep-H [43]; HeLa - H1-H [68]; COS - H1-H (our unpublished data); S.c. - DPAPB-H [49]; S.c. - CPY-H [52]; S.c. - SP-Lep-H [69]; E.c. - PCLep-H [70]; E.c. - Lep<sup>LacY</sup>-H [71]. S.c., *Saccharomyces cerevisiae*; E.c., *Escherichia coli*. The model constructs used are shown schematically with signal and TM segments colored according to their function (as in Fig. 2): red for cleavable signals and signal-anchors, turquoise for reverse signal anchors, pink for re-integration sequences, and blue for the H-segments (potential stop-transfer sequences) with the GPGG...GGPG insulator sequences in yellow. The sequence lines are drawn in different grays to indicate their different origins. A scale is shown for polypeptide length in amino acids (aa) and total protein lengths are provided in parentheses. Black dots indicate glycosylation sites and arrowheads signal cleavage sites.