FEBS 25121 FEBS Letters 504 (2001) 87–93

Minireview

Topogenesis of membrane proteins: determinants and dynamics

Veit Goder, Martin Spiess*

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

Received 11 June 2001; accepted 26 June 2001

First published online 25 July 2001

Edited by Andreas Engel and Giorgio Semenza

Abstract For targeting and integration of proteins into the mammalian endoplasmic reticulum, two types of signals can be distinguished: those that translocate their C-terminal sequence (cleavable signals and signal-anchors) and those that translocate their N-terminus (reverse signal-anchors). In addition to the well established effect of flanking charges, also the length and hydrophobicity of the apolar core of the signal as well as protein folding and glycosylation contribute to orienting the signal in the translocon. In multi-spanning membrane proteins, topogenic determinants are distributed throughout the sequence and may even compete with each other. During topogenesis, segments of up to 60 residues may move back and forth through the translocon, emphasizing unexpected dynamic aspects of topogenesis.

Key words: Endoplasmic reticulum; Glycosylation; Signal recognition particle; Signal sequence

1. Signals translocate either their C- or N-terminus

In mammalian cells, protein targeting to the endoplasmic reticulum (ER) is mostly cotranslational [1], i.e. protein translocation and membrane insertion are coupled to protein synthesis. As a hydrophobic signal sequence emerges from the ribosome, it is recognized by a signal recognition particle (SRP) in the context of the nascent chain–ribosome complex. This complex is targeted to the ER membrane by binding to the SRP receptor (SR). The signal sequence then interacts with the Sec61α subunit of the translocon [2,3] and initiates translocation of the polypeptide. Synchronization of translation with membrane targeting and insertion prevents the protein from premature folding or aggregation in the cytosol. This applies particularly to membrane proteins: even in systems where posttranslational targeting and translocation of secretory proteins are common, as in *Escherichia coli* or yeast, most membrane proteins and particularly the multi-spanning ones are targeted to the membrane in an SRP-dependent manner and are thus inserted cotranslationally [4].

The mammalian translocon consists of three to four copies of the heterotrimeric Sec61 αβγ complex forming a gated pore [5]. By cross-linking studies, TRAM (translocating chain-associating membrane protein) was shown to be in contact with signal sequences and transmembrane segments early during translocation and to be required for efficient insertion of at least some substrate proteins. In addition, homologues of the yeast Sec62/63 complex have been identified [6,7]. By its J-domain, Sec63 recruits the lumenal chaperone BiP to the translocon where it may contribute to sealing the pore and/or translocation [8]. This machinery is responsible for the translocation of secretory proteins and the membrane insertion of the vast majority of membrane proteins.

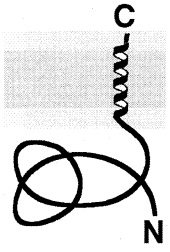
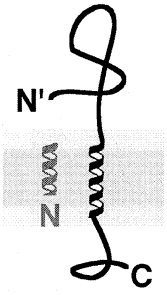
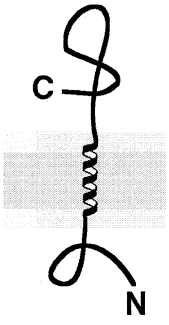
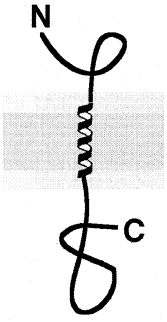
Single-spanning membrane proteins may assume a final topology with a cytoplasmic N- and an exoplasmic C-terminus (N_{cyt}/C_{exo}) or with the opposite orientation (N_{exo}/C_{cyt}). However, if the mechanism of insertion is taken into consideration, four major types of single-spanning membrane proteins can be distinguished, as is summarized in Table 1. Type I membrane proteins are initially targeted to the ER by an N-terminal, *cleavable signal sequence*, a hydrophobic stretch of typically 7–15 predominantly apolar residues, and then anchored in the membrane by a subsequent *stop-transfer sequence*, a segment of ~20 hydrophobic residues that halts the further translocation of the polypeptide and acts as a transmembrane anchor. In type II membrane proteins, a *signal-anchor sequence* is responsible for both insertion and anchoring. Signal-anchor sequences are generally longer than cleaved signals (~18–25 mostly apolar amino acids), since they span the lipid bilayer as a transmembrane helix. They lack a signal peptidase cleavage site and they can be positioned internally within the polypeptide chain. However, like cleaved signals, they induce the translocation of their C-terminal end across the membrane. The opposite is the case for *reverse signal-anchors* of type III proteins, which translocate their N-terminal end across the membrane. These three types of membrane proteins are all inserted by the same machinery involving SRP, SRP receptor and the Sec61 translocon [9]. With respect to topogenesis, there are two basic types of signals translocating either their C-terminus (cleaved signals and signal-anchors) or their N-terminus (reverse signal-anchors), illustrated in Fig. 1.

In addition, there is also a class of proteins predominantly exposed to the cytosol and anchored to the membrane by a very C-terminal signal sequence. Examples are cytochrome b_5 and the SNARE proteins like synaptobrevin. Insertion of these proteins is necessarily posttranslational, since the signal emerges from the ribosome only after translation has reached the stop codon. Accordingly, targeting and insertion was found to be independent of SRP and the Sec61 complex, and to use an as yet unknown ATP-requiring mechanism [10–12]. This last group of proteins will not be further considered here.

*Corresponding author. Fax: (41)-61-267 2149.
E-mail address: martin.spiess@unibas.ch (M. Spiess).

PII: S0014-5793(01)02712-0

Table 1
Topogenic determinants of single-spanning membrane proteins

Signal type:	C-Terminus-translocating signals			N-Terminus-translocating
Topogenic determinants:	C-terminal signal	Cleaved signal + stop transfer sequence	Signal-anchor	Reverse signal-anchor
Machinery:	unknown	SRP/SR/Sec61 + signal peptidase	SRP/SR/Sec61	SRP/SR/Sec61
exoplasmic / Final topology: / cytoplasmic		 Type I	 Type II	 Type III
Examples:	Synaptobrevin; cytochrome b_5	Glycophorin; LDL receptor	Transferrin receptor; galactosyl transferase	Synaptotagmin; neuregulin; cytochromes P-450

2. Multiple determinants of signal sequence orientation in the membrane

2.1. Flanking charges

The best established determinant of signal orientation (and also so far the most useful one for topology prediction) is the distribution of charged residues near the hydrophobic core of signal and transmembrane segments. The 'positive-inside rule' was first discovered for bacterial proteins where positive residues were statistically found to be four times more abundant in cytoplasmic than in periplasmic loops [13]. An opposite, but weaker correlation exists for acidic amino acids. A similar charge bias was also observed for membrane proteins in the ER, chloroplasts and mitochondria [14,15]. For ER signals, the charge difference between the two flanking segments of a signal's hydrophobic core, rather than the positive charge per se correlates with transmembrane orientation: the more positive flanking sequence is generally cytoplasmic [16]. Experimentally, a type III cytochrome *P*-450 could be converted to a type II protein by insertion of positively charged residues into its short N-terminal domain [17–19]. Mutation of flanking charges in the asialoglycoprotein (ASGP) receptor H1 and in the paramyxovirus hemagglutinin neuraminidase, two type II proteins, caused a fraction of the polypeptides to insert with the opposite type III topology (Fig. 2A) [20–22]. However, since in these as well as other studies (e.g. [23,24]) the asymmetric distribution of flanking charges in mutant proteins was not sufficient to generate a unique topology, it is clear that additional factors must contribute to uniform topogenesis.

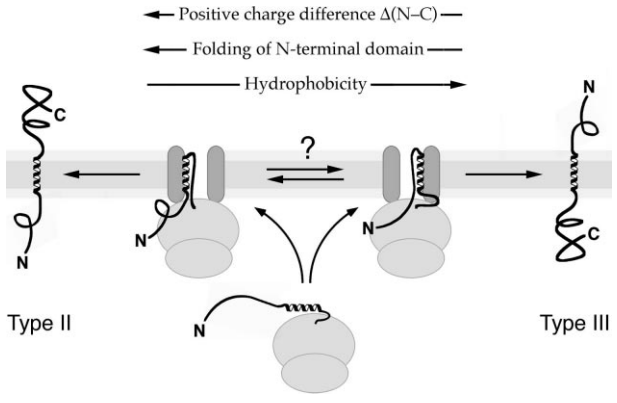


Fig. 1. Signals translocate either their C- or their N-terminus. Signal-anchor and reverse signal anchor sequences engage with the translocon in opposite orientations resulting in type II or type III membrane proteins, respectively. Cleavable signals behave like signal-anchors until cleavage releases the C-terminal domain into the ER lumen (not indicated). The know determinants of signal orientation are listed above. The signals may reorient within the translocon.

Charged residues near the hydrophobic core of a signal are likely to exert their topogenic effect by interaction with negative charges at the cytoplasmic and/or positive charges at the exoplasmic sides of the translocon. As yet, no relevant charged residues have been identified in the proteins of the translocation machinery. In the bacterial system, there is evidence for an electrostatic interaction of positive flanking charges with negative lipids on the cytosolic membrane surface, resulting in retention of the positive portion on the cytoplasmic face of the membrane [25,26]. An interaction of the

flanking sequences with lipids in or near the translocon cannot be excluded for the eukaryotic system either.

2.2. Folding of the N-terminal domain

Cotranslational translocation ensures the direct transfer of sequences downstream of cleavable signals and signal-anchors across the membrane and prevents exposition to and folding in the cytosol. (An exception is the 'paused' translocation shown in translocation of apolipoprotein B which allowed sequences to loop into the cytosol [27].) In contrast, sequences N-terminal to an internal signal are exposed to the cytosol before the targeting signal emerges from the ribosome. Folding of these domains may thus affect their translocation competence and favor retention of the N-terminus in the cytosol. This was indeed demonstrated using mutants of the ASGP receptor (Fig. 2). The ASGP receptor is a type II membrane protein with a typical charge distribution: two N-terminal arginines and two C-terminal aspartates flank the signal sequence. Inversion of these charges yielded equal fractions with each orientation (construct A [20]). Additional truncation of the N-terminal domain allowed almost complete N-terminal translocation (B), whereas the full coding sequence of dihydrofolate reductase (C) or a small zinc finger domain (E) fused at the N-terminus hindered or even blocked it [28]. This was due to protein folding and not simply the size of the N-terminal extensions, since disruption of their structure by destabilizing point mutations (D and F) largely recovered N-terminal transport across the membrane. These results confirmed that the polypeptide chain needs to be unfolded for translocation and that the folding properties of the N-terminal domain influence protein orientation.

Many type III proteins lack a sizeable N-terminal domain, facilitating transfer of the N-terminus. For example, most members of the cytochrome *P*-450 family have just a few N-terminal residues to translocate. In contrast, NTAK protein (neural- and thymus-derived activator for ErbB kinases) and the neuregulin precursor translocate N-terminal sequences of 405 and 241 residues, respectively, including immunoglobulin-like and EGF-like domains. These sequences apparently do not stably fold during the time they are exposed to the cytosol. They are likely to be associated with cytosolic chaperones until translocation. In addition, disulfide formation, which stabilizes these domains, will only occur when they have reached the ER lumen.

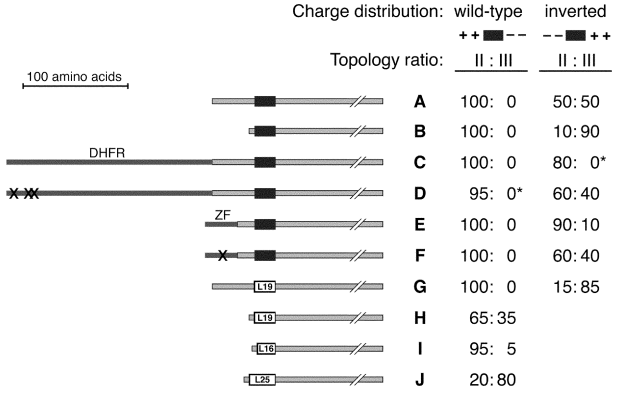


Fig. 2. Multiple determinants define the orientation of ASGP receptor mutants. The sequence of the ASGP receptor and various derivatives are shown schematically, with a black box indicating the hydrophobic core of the internal signal sequence. DHFR and ZF denote the fused sequences of mouse dihydrofolate reductase and the yeast zinc finger domain ADR1a. Point mutations affecting protein folding are indicated by Xs. The black box represents the hydrophoic core of the signal-anchor sequence, the white box generic oligo-leucine sequences of 19 (G, H), 16 (I), or 25 leucines (J). The constructs were expressed in vivo in transfected COS cells. The ratio of type II:type III insertion is listed for the constructs with the wild-type distribution of charges flanking the signal-anchor (positive Δ(N−C), left column) and with inverted flanking charges (negative Δ(N−C), right column). *A fraction of the products (the difference to 100%) was not inserted into the membrane. References: A, B [20]; C–F [28]; G, H [33].

2.3. Hydrophobicity

A topogenic contribution of the hydrophobic segment of signal and signal-anchor sequences was first suggested by in vitro experiments by Omura and colleagues. Short deletions within the hydrophobic segment of the N_{exo}/C_{cyt} signal-anchor sequence of a cytochrome *P*-450 resulted partially in translocation of the C-terminal reporter sequence [29]. Similarly, an artificial signal sequence of 12 or fewer leucines with a negative N-terminal net charge was found to translocate its C-terminal reporter sequence. However, with longer hydrophobic segments of 13 or 15 leucines a fraction of the polypeptides was anchored in microsomal membranes as type III proteins [30]. Mutations which extend the hydrophobic core of the cleaved signal of the pseudorabies virus glycoprotein gC were observed to reduce C-terminal translocation, suggesting that there is a limit to hydrophobicity of a cleavable export signal [31]. Likewise, artificial signals with hydrophobic cores of 5–20 leucine residues were found to be recognized with increasing affinity by SRP (as judged by SRP-induced translation arrest in the absence of microsomal membranes). Yet, translocation of the downstream sequence decreased with more than 14 leucines [32]. The possibility of membrane insertion in an N_{exo}/C_{cyt} orientation, however, had not been tested in either study.

Systematic analysis of the effects of different hydrophobic domains on orienting a signal sequence in vivo was performed using the ASGP receptor as a model protein. It was initially observed that the receptor with a truncated N-terminal domain and with a generic Leu_{19} sequence replacing the natural 19-residue hydrophobic core of the signal-anchor (Fig. 2, construct H) did not translocate the C-terminus in one third of the products [33]. These polypeptides were shown to be inserted as type III proteins, since alkaline and mild saponin extraction did not release them from the membrane, and protease protection assays confirmed the cytoplasmic exposition of the C-terminus. Extending the oligo-leucine sequence up to 25 residues and thus increasing the total hydrophobicity gradually increased the fraction of N_{exo}/C_{cyt} proteins to $\sim$80% with Leu_{25}, whereas stretches of 16 or less leucines almost exclusively yielded type II proteins (constructs J and I, respectively).

N-Terminal translocation was induced by long oligo-leucine sequences despite a typical type II charge distribution with two positive flanking charges at the N- and two negative ones at the C-terminus. Reduction of the N-terminal charge from +2 to +1 facilitated N-terminal translocation. As a result, the oligo-leucine series covered the entire spectrum from almost complete C-terminal translocation with Leu_7 to exclusive N-terminal translocation with Leu_{22} and Leu_{25}. In the presence of the wild-type N-terminal domain of 40 residues, oligo-leucine sequences did not alter type II insertion, except

in combination with inverted flanking charges (Fig. 2, construct G). The influence of the hydrophobic sequence on topology was thus additive with the effects of flanking charges and of an N-terminal hydrophilic extension [33,34]. This may explain how those natural signal sequences which violate the charge rules succeed in acquiring their correct and uniform orientation. Experimentally, this could be shown by extending the apolar sequence of the cleavable signal of the vasopressin precursor ($\Delta(N-C) = -3$) and by shortening that of the reverse signal-anchor of microsomal epoxide hydrolase ($\Delta(N-C) = +3$). In both cases, a significant fraction of mutant polypeptides inserted in the opposite orientation to that of the wild-type proteins, demonstrating the physiological significance of the topogenic contribution of the hydrophobic segment [35].

It is not trivial to separate the potential topogenic effects of hydrophobicity, length, and other properties of the signal core, such as helix propensity and shape. Homo-oligomers of other apolar sequences than oligo-leucines showed a similar bias for increasing N-terminal translocation with increasing length [36]. For example Val_{16} to Val_{25} sequences in the context of ASGP receptor without N-terminal domain and with an N-terminal charge of +1 covered the spectrum from unique N_{cyt}/C_{exo} to exclusive N_{exo}/C_{cyt} insertion. The ability of different homo-oligomers to promote N-terminal translocation decreased in the order $I > L > V \sim W > Y > F > M$. Except for oligo-alanine, which was not functional as a signal sequence, all homo-oligomers tested were efficient in targeting and insertion, highlighting the ability of both SRP and the translocon to accommodate an extremely broad spectrum of signal sequences. Sequences as different in shape and volume as Val_{19} and Trp_{19} (Fig. 3) behaved even identically with respect to topogenesis.

In a more natural setting, all uncharged amino acids have also been tested for their effect on signal orientation in a host–guest approach: two 'guest residues' of each type were inserted into an oligo-leucine 'host sequence' with a total of 16 or 19 residues in the context of N-terminally truncated ASGP receptor [36]. Leucine is the most abundant amino acid in transmembrane and signal sequences, accounting for ~25% of the residues. The ranking order of the amino acids with respect to promoting N-terminal translocation was similar to that of the homo-oligomers: $I > V > L \sim W > F > Y > C > M > A > T > S > G > N > Q > H > P$. This ranking resembles a hydrophobicity scale, but also a scale of helix propensities in an apolar environment [37]. Harley et al. [34] also observed a correlation between signal orientation and a hydropathy gradient along the apolar sequence: the more hydrophobic end was more efficiently translocated across the membrane.

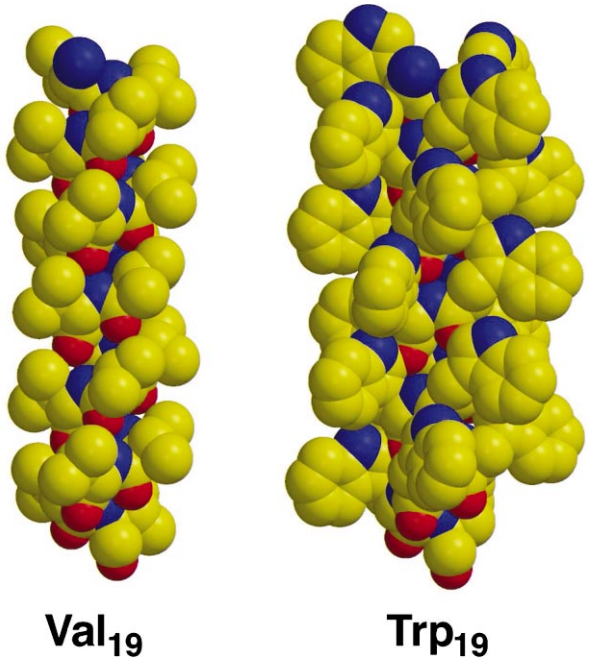


Fig. 3. Models of Val_{19} and Trp_{19} α-helices. Both sequences efficiently target the nascent protein to the ER and behave identically with respect to topogenesis in vivo [36].

2.4. A hypothesis for the hydrophobic contribution to topogenesis

The signal is first bound to SRP in a hydrophobic groove formed by the methionine-rich domain of subunit SRP54 [38], then it is recognized by the Sec61 translocation complex [39]. By photocross-linking, it was found in contact with Sec61α and lipids [3,40]. The hydrophobic segment of the signal was found to specifically associate with transmembrane helices 2 and 7 of Sec61p, the yeast homologue of Sec61α [41]. In addition, TRAM could also be cross-linked to signal sequences [2,42–44].

How do the properties of the hydrophobic domain influence topogenesis? The effect of a hydrophobicity gradient in the signal could be explained by a similar gradient in the signal binding site of the translocon. However, it is more difficult to rationalize how uniform sequences like Leu_{19}, Val_{19}, and Phe_{19} select different preferred orientations. In the apolar binding sites of SRP and of the translocon, the hydrophobic core of the signal is likely to assume a helical conformation. It is not known how the signal is transferred from the first binding site to the second, except that GTP is hydrolyzed in both SRP and SR [45–47]. An attractive model would not require the hydrophobic signal to completely dissociate from SRP into the aqueous milieu before rebinding to the interaction site in the translocon. Instead, docking of SRP to its receptor might optimally position it relative to the translocon to allow the signal to shift into the translocon site on a 'greasy slide' (possibly facilitated by a conformational change in SRP54 upon GTP hydrolysis). The signal will thus insert head-on, generating an N_{exo}/C_{cyt} orientation (Fig. 4, right branch). Yet, a positive N-terminal flanking sequence will interact with negative charges at the cytosolic surface of the translocon, either on proteins or on lipids, and tend to anchor the N-terminus of the signal on the cytosolic side and force the signal to invert its orientation and form a translocation loop (Fig. 4, left branch). This might occur in a hydrophobic cleft between translocon subunits and involve lipids, in agreement with cross-linking data, or it might require partial dissociation into the polar environment of the translocon pore. In either case, it is plausible that short, less hydrophobic, or unstable helices will reorient more readily in response to a positive charge difference $\Delta(N-C)$ than long, more hydrophobic, and stable helices. Similarly, an N-terminal extension, particularly if already folded, will favor or require inversion of the signal to an N_{cyt}/C_{exo} orientation.

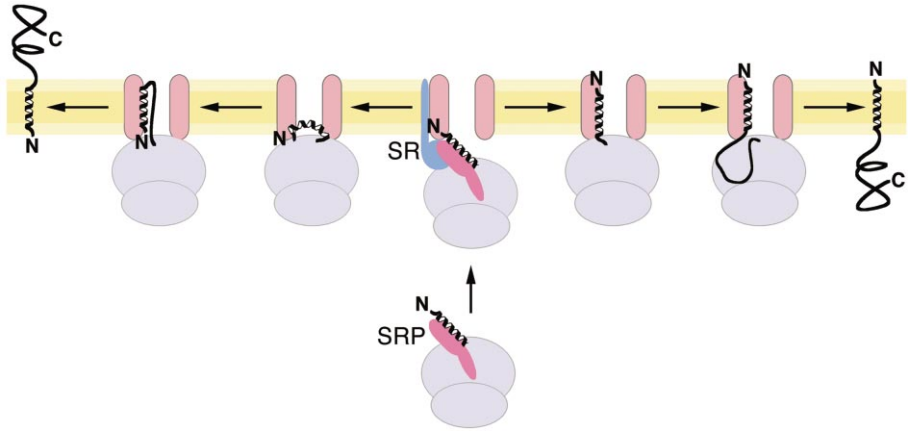


Fig. 4. Model for the topogenic effect of the hydrophobic core of the signal. For explanation see text.

2.5. Kinetics of targeting and translocation

SRP binding to the signal sequence was shown in vitro to reduce the elongation rate of the nascent chain–ribosome complex until its productive docking to the ER membrane [48–50]. Evidence for a physiological role of this inhibition has recently been reported in yeast expressing an SRP mutant defective in slowing elongation [50]. Tight coupling of the protein's translation to its transfer through the translocon ensures translocation competence of the sequences C-terminal of the signal. In contrast, N-terminal domains of type III proteins are already completed by the time the signal emerges from the ribosome and are at risk to fold or misfold still in the cytosol. The kinetics of SRP binding, targeting to the ER membrane, and translocation are thus particularly relevant for type III proteins.

A model type III protein with an N-terminal consensus sequence for phosphorylation by protein kinase A was used to determine these kinetics in vivo in transfected COS cells [51]. Phosphorylation and thus labeling with [^{32}P]phosphate is restricted to the time the phosphorylation site is exposed to the cytosol: from when it emerges from the ribosome until the signal does, plus the time it takes to bind SRP to the signal and to target the complex to the ER membrane, plus the time required to translocate the N-terminal domain across the membrane. Increasing the time of cytosolic exposure by extending the N-terminal domain or by reducing the rate of translation with cycloheximide increased the specific phosphorylation per newly synthesized protein, indicating that the specific phosphorylation acts as a molecular timer. For the model protein used (with a hydrophobic core of 25 leucine residues), the average time of SRP binding and targeting was determined to amount to only a few seconds (2–5 s). At a normal elongation rate of ~5 amino acids/s [52], this amounts to a maximum (i.e. in the absence of any translation slowdown) of 10–25 residues. This does not exclude an important role for an SRP-induced translation slowdown, since targeting of the first ribosome on an mRNA may take significantly longer than that of the subsequent ribosomes in a polysome that is already tethered to the ER membrane. Furthermore, these experiments allowed to estimate the rate of translocation of the N-terminal sequence to be ~8 amino acids/s, i.e. approx. 1.6 times the rate of translation.

3. Topogenesis of multi-spanning membrane proteins

3.1. Topogenic determinants are present throughout the polypeptide

For complex membrane proteins, which span the bilayer multiple times, it is generally assumed that the first hydrophobic sequence is responsible for targeting the nascent protein to the ER and to initiate translocation and membrane insertion. Accordingly, these proteins could be classified as multi-spanning proteins of type I, II, or III based on whether the most N-terminal apolar sequence is cleaved by signal peptidase or spans the membrane with an N_{cyt}/C_{exo} or N_{exo}/C_{cyt} orientation, respectively [53]. The superfamily of seven-transmembrane receptors, for example, consists of ~15% members with a cleavable signal to translocate what becomes the N-terminal exoplasmic domain. The other ones are inserted with a reverse signal-anchor as type III seven-spanning proteins. (Interestingly, the average size of the N-terminal domain is ~140 residues for the type I members and only ~40 amino

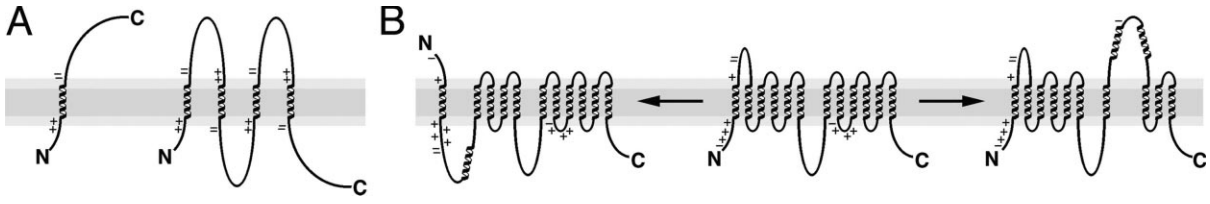


Fig. 5. Topogenic determinants can be functional throughout the polypeptide. A: Proteins may insert linearly overriding downstream determinants, if the spacer sequences are sufficiently long; e.g. the signal-anchor sequence of the ASGP receptor (left), when repeated four times in a model polypeptide separated by ~100 residues, is inserted with either N_{cyt}/C_{exo} or N_{exo}/C_{cyt} orientation depending on its relative position in the protein [55]. B: In natural proteins with short loops between transmembrane segments, mutation of topogenic determinants disturbs topology only locally; e.g. charge mutations in Glut1 in the flanking sequences of the first transmembrane domain (left) [60] or in the eighth loop (right) [61] only affect insertion of the neighboring transmembrane segments.

acids for the type III members, probably reflecting the folding-related difficulties in translocating large N-terminal sequences.)

According to the simplest model, the most N-terminal signal sequence defines its own orientation as well as the orientations of all subsequent transmembrane segments, which will act alternately as stop-transfer and reinsertion sequences. They do not require any additional information, but will simply follow the lead of the first signal. Evidence for this 'linear insertion model' (initially proposed by Blobel [54]) has been provided using chimeric proteins with two to four signal-anchor sequences separated by 50–200 residues from each other [55,56]. The even-numbered signal-anchors (normally N_{cyt}/C_{exo}) were indeed forced to insert as stop-transfer sequences (N_{exo}/C_{cyt}; Fig. 5A).

However, there is also strong evidence against a dominant role of the first signal in many membrane proteins. Statistically, internal transmembrane domains also follow the charge rules, although less stringently than the most N-terminal ones [57]. Experimentally, insertion of clusters of positive charges into short exoplasmic loops of model proteins caused individual hydrophobic domains not to insert at all ('frustrated' topologies) [58], showing that positive charges inside a polytopic membrane protein can be topogenically active. In bacterial lactose permease and MalF, deletion of individual membrane-spanning segments did not affect the topology of the downstream transmembrane domains [59]. For the human glucose transporter Glut1 with 12 membrane spans, inversion of the charge difference of the first one did not alter the topology of the rest of the molecule, but prevented insertion of the second apolar sequence (Fig. 5B, left) [60]. Mutation of the positive charges in the conserved sequence RXGRR in the first or fourth cytosolic loop of Glut1 resulted only in a local perturbation in the membrane topology in which the cytoplasmic loop was aberrantly translocated into the exoplasm along with the two flanking transmembrane segments (Fig. 5B, right) [61]. These studies showed that multi-spanning proteins contain functional topogenic information throughout their sequence.

3.2. Dynamic reorientation of nascent polypeptides in the translocon

There is an apparent discrepancy between the early studies with artificial model proteins, which supported a dominant role of the most N-terminal signal [55,56], and those with mostly natural proteins, indicating that the topology is defined by multiple dispersed determinants [58–62]. This may be explained by the relatively long spacers between conflicting determinants used in the former studies in comparison to the frequently short loops connecting transmembrane segments. This was indeed confirmed by a systematic analysis using simple chimeric proteins with two conflicting signal sequences, a cleavable signal and a signal-anchor (Fig. 6A) [63]. When the signals were separated by 80 residues or more, linear insertion overriding the topological preference of the second signal was observed (Fig. 6A, left topology). With shorter spacers, however, an increasing fraction of proteins inserted with a translocated C-terminus as dictated by the second signal (Fig. 6A, right topology). The second signal thus co-determined the insertion process. At an elongation rate of ~5 amino acids/s, the second signal will enter the translocon approx. 16 s after the first one (60 residues spacer+20 residues of the signal-anchor, divided by 5 amino acids/s). This indicates that at least a subset of polypeptides is not committed to a final topology 16 s after initiation of topogenesis.

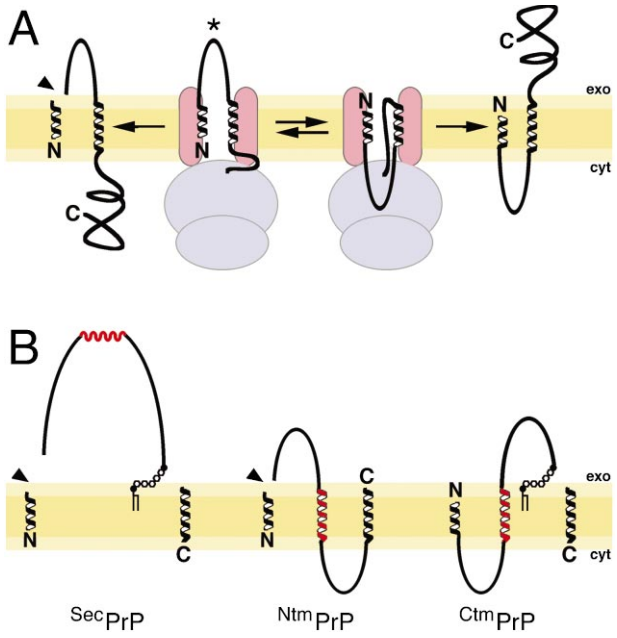


Fig. 6. Competition between topogenic determinants. A: Combination of a cleavable signal with a signal-anchor sequence generated two different topologies [63]. Spacer sequences of 80 or more residues produced exclusively the type I membrane protein on the left. Shorter spacers yielded an increasing fraction of the topology on the right. Glycosylation in the loop (indicated by an asterisk) shifted the equilibrium to the left. B: Topologies of the prion protein. Signal cleavage is indicated by an arrowhead.

In most natural proteins, the contributions of subsequent determinants add up to a unique topology. Yet, the human prion protein is an example in which competing signals appear to be at work similar to the artificial constructs mentioned above (Fig. 6B) [64]. The prion protein contains a cleavable signal at the N-terminus and a stop-transfer sequence which is normally replaced by a glycosylphosphatidylinositol (GPI)-anchor at the very C-terminus. In addition there is a hydrophobic sequence in the middle of the protein which is mostly translocated across the membrane, generating the so-called ^{Sec}PrP. In a fraction of the products, however, it is integrated into the bilayer either with an N_{exo}/C_{cyt} orientation, as expected for a stop-transfer sequence, producing ^{Ntm}PrP, or with the opposite N_{cyt}/C_{exo} orientation, producing ^{Ctm}PrP. Natural and artificial mutants in or near the internal hydrophobic sequence and the N-terminal signal are known which increase the fraction of ^{Ctm}PrP [64].

It is obvious that two hydrophobic sequences with a very short spacer cannot orient themselves independently of each other during topogenesis, since they will insert together as a hairpin. With an increasing distance the cooperativity (or the competition) of successive transmembrane elements decreases and finally disappears. In the case of contradictory determinants, the polypeptide segment translocated by the first signal will have to be retranslocated to the cytosolic side when the second signal enters the translocon and exerts its topogenic preference. This was confirmed by insertion of an *N*-glycosylation site into the spacer sequence between the two conflicting signals illustrated in Fig. 6A [63]. Glycosylation significantly

shifted the equilibrium of topologies in favor of that with a translocated spacer sequence, most likely by sterically trapping it in the ER lumen. This demonstrates that polypeptides of up to 60 residues in length can dynamically reorient within the translocon and that glycosylation can co-determine protein topology.

4. Conclusion

Protein topogenesis appears to be directed by multiple determinants which in unlimited combinations add up to an almost uniform result for most, but not all, natural proteins. With what is known up to now, it is still impossible, certainly in the case of complex membrane proteins, to confidently predict topology from sequence or to design membrane proteins de novo. During protein targeting, translocation, and insertion into the membrane, several processes occur simultaneously and influence each other. Modifications like glycosylation, but potentially also signal cleavage and protein folding, may stabilize particular topologies. Furthermore, relatively large segments can move back and forth through the translocation pore in search for the most comfortable position of the polypeptide in the translocon and the membrane. Protein insertion is thus in unexpected ways a very dynamic process.

Acknowledgements: We are grateful to Dr. Peter Burkhard for Fig. 3 and fruitful discussions. Our own work was supported by Grant 31-61579.00 from the Swiss National Science Foundation.

References

[1] Walter, P. and Johnson, A.E. (1994) Annu. Rev. Cell Biol. 10, 87–119.
[2] Mothes, W., Prehn, S. and Rapoport, T.A. (1994) EMBO J. 13, 3973–3982.
[3] Mothes, W., Jungnickel, B., Brunner, J. and Rapoport, T.A. (1998) J. Cell Biol. 142, 355–364.
[4] Valent, Q.A., Scotti, P.A., High, S., de Gier, J.W., von Heijne, G., Lentzen, G., Wintermeyer, W., Oudega, B. and Luirink, J. (1998) EMBO J. 17, 2504–2512.
[5] Johnson, A.E. and van Waes, M.A. (1999) Annu. Rev. Cell Dev. Biol. 15, 799–842.
[6] Tyedmers, J., Lerner, M., Bies, C., Dudek, J., Skowronek, M.H., Haas, I.G., Heim, N., Nastainczyk, W., Volkmer, J. and Zimmermann, R. (2000) Proc. Natl. Acad. Sci. USA 97, 7214–7219.
[7] Meyer, H.A., Grau, H., Kraft, R., Kostka, S., Prehn, S., Kalies, K.U. and Hartmann, E. (2000) J. Biol. Chem. 275, 14550–14557.
[8] Young, B.P., Craven, R.A., Reid, P.J., Willer, M. and Stirling, C.J. (2001) EMBO J. 20, 262–271.
[9] High, S. and Dobberstein, B. (1991) J. Cell Biol. 113, 229–233.
[10] Kutay, U., Hartmann, E. and Rapoport, T.A. (1993) Trends Cell Biol. 3, 72–75.
[11] Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B. and Rapoport, T.A. (1995) EMBO J. 14, 217–223.
[12] Whitley, P., Grahn, E., Kutay, U., Rapoport, T.A. and von Heijne, G. (1996) J. Biol. Chem. 271, 7583–7586.
[13] von Heijne, G. (1986) EMBO J. 5, 3021–3027.
[14] von Heijne, G. (1992) J. Mol. Biol. 225, 487–494.
[15] Wallin, E. and von Heijne, G. (1998) Protein Sci. 7, 1029–1038.
[16] Hartmann, E., Rapoport, T.A. and Lodish, H.F. (1989) Proc. Natl. Acad. Sci. USA 86, 5786–5790.
[17] Monier, S., Van Luc, P., Kreibich, G., Sabatini, D.D. and Adesnik, M. (1988) J. Cell Biol. 107, 457–470.
[18] Szczesna-Skorupa, E., Browne, N., Mead, D.A. and Kemper, B. (1988) Proc. Natl. Acad. Sci. USA 85, 738–742.
[19] Szczesna-Skorupa, E. and Kemper, B. (1989) J. Cell Biol. 108, 1237–1243.
[20] Beltzer, J.P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H.P. and Spiess, M. (1991) J. Biol. Chem. 266, 973–978.
[21] Parks, G.D. and Lamb, R.A. (1991) Cell 64, 777–787.
[22] Parks, G.D. and Lamb, R.A. (1993) J. Biol. Chem. 268, 19101–19109.
[23] Andrews, D.W., Young, J.C., Mirels, L.F. and Czarnota, G.J. (1992) J. Biol. Chem. 267, 7761–7769.
[24] Harley, C.A. and Tipper, D.J. (1996) J. Biol. Chem. 271, 24625–24633.
[25] Gallusser, A. and Kuhn, A. (1990) EMBO J. 9, 2723–2729.
[26] van Klompenburg, W., Nilsson, I., von Heijne, G. and de Kruijff, B. (1997) EMBO J. 16, 4261–4266.
[27] Hegde, R.S. and Lingappa, V.R. (1996) Cell 85, 217–228.
[28] Denzer, A.J., Nabholz, C.E. and Spiess, M. (1995) EMBO J. 14, 6311–6317.
[29] Sato, T., Sakaguchi, M., Mihara, K. and Omura, T. (1990) EMBO J. 9, 2391–2397.
[30] Sakaguchi, M., Tomiyoshi, R., Kuroiwa, T., Mihara, K. and Omura, T. (1992) Proc. Natl. Acad. Sci. USA 89, 16–19.
[31] Tomilo, M., Wilkinson, K.S. and Ryan, P. (1994) J. Biol. Chem. 269, 32016–32021.
[32] Hatsuzawa, K., Tagaya, M. and Mizushima, S. (1997) J. Biochem. 121, 270–277.
[33] Wahlberg, J.M. and Spiess, M. (1997) J. Cell Biol. 137, 555–562.
[34] Harley, C.A., Holt, J.A., Turner, R. and Tipper, D.J. (1998) J. Biol. Chem. 273, 24963–24971.
[35] Eusebio, A., Friedberg, T. and Spiess, M. (1998) Exp. Cell Res. 241, 181–185.
[36] Rösch, K., Naeher, D., Laird, V., Goder, V. and Spiess, M. (2000) J. Biol. Chem. 275, 14916–14922.
[37] Liu, L.P. and Deber, C.M. (1998) J. Biol. Chem. 273, 23645–23648.
[38] Bui, N. and Strub, K. (1999) Biol. Chem. 380, 135–145.
[39] Jungnickel, B. and Rapoport, T.A. (1995) Cell 82, 261–270.
[40] Martoglio, B., Hofmann, M.W., Brunner, J. and Dobberstein, B. (1995) Cell 81, 207–214.
[41] Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J. and Rapoport, T.A. (1998) Cell 94, 795–807.
[42] Görlich, D. and Rapoport, T.A. (1993) Cell 75, 615–630.
[43] High, S., Martoglio, B., Gorlich, D., Andersen, S.S.L., Ashford, A.J., Giner, A., Hartmann, E., Prehn, S., Rapoport, T.A., Dobberstein, B. and Brunner, J. (1993) J. Biol. Chem. 268, 26745–26751.
[44] Hegde, R.S., Voigt, S., Rapoport, T.A. and Lingappa, V.R. (1998) Cell 92, 621–631.
[45] Rapiejko, P.J. and Gilmore, R. (1997) Cell 89, 703–713.
[46] Fulga, T.A., Sinning, I., Dobberstein, B. and Pool, M.R. (2001) EMBO J. 20, 2338–2347.
[47] Song, W., Raden, D., Mandon, E.C. and Gilmore, R. (2000) Cell 100, 333–343.
[48] Wolin, S.L. and Walter, P. (1989) J. Cell Biol. 109, 2617–2622.
[49] Ogg, S.C. and Walter, P. (1995) Cell 81, 1075–1084.
[50] Mason, N., Ciufo, L.F. and Brown, J.D. (2000) EMBO J. 19, 4164–4174.
[51] Goder, V., Crottet, P. and Spiess, M. (2000) EMBO J. 19, 6704–6712.
[52] Hershey, J.W.B. (1991) Annu. Rev. Biochem. 60, 717–755.
[53] Spiess, M. (1995) FEBS Lett. 369, 76–79.
[54] Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 1496–1500.
[55] Wessels, H.P. and Spiess, M. (1988) Cell 55, 61–70.
[56] Lipp, J., Flint, N., Haeuptle, M.T. and Dobberstein, B. (1989) J. Cell Biol. 109, 2013–2022.
[57] von Heijne, G. (1989) Nature 341, 456–458.
[58] Gafvelin, G., Sakaguchi, M., Andersson, H. and von Heijne, G. (1997) J. Biol. Chem. 272, 6119–6127.
[59] McGovern, K. and Beckwith, J. (1991) J. Biol. Chem. 266, 20870–20876.
[60] Sato, M., Hresko, R. and Mueckler, M. (1998) J. Biol. Chem. 273, 25203–25208.
[61] Sato, M. and Mueckler, M. (1999) J. Biol. Chem. 274, 24721–24725.
[62] Bibi, E., Verner, G., Chang, C.Y. and Kaback, H.R. (1991) Proc. Natl. Acad. Sci. USA 88, 7271–7275.
[63] Goder, V., Bieri, C. and Spiess, M. (1999) J. Cell Biol. 147, 257–266.
[64] Stewart, R.S., Drisaldi, B. and Harris, D.A. (2001) Mol. Biol. Cell 12, 881–889.