Membrane protein integration into the ER

Journal:	*FEBS Journal*
Manuscript ID:	FJ-11-0304.R1
Manuscript Type:	Invited Review
Subdiscipline:	Membranes
Date Submitted by the Author:	n/a
Complete List of Authors:	Martinez-Gil, Luis; University of Valencia, Biochemistry and Molecular Biology Saurí, Ana; VU University, Molecular Microbiology Marti-Renom, Marc; CIPF: Centre of Research Príncipe Felipe, Bioinformatics and Genomics Dept. Mingarro, Ismael; University of Valencia, Dept. of Biochemistry & Molecular Biology
Key Words:	biogenesis, insertion, membrane protein, translocon, transmembrane segment

13 May 2011

Membrane protein integration into the ER

Luis Martínez-Gil [1,2=], Ana Saurí [3=], Marc A. Marti-Renom [4] and Ismael Mingarro [1] *

[1] Departament de Bioquímica i Biologia Molecular, Universitat de València. E46100 Burjassot, Spain.

[2] Department of Microbiology, Mount Sinai School of Medicine, One Gustave Levy Place, New York, NY 10029-6574, USA.

[3] Department of Molecular Microbiology, Institute of Molecular Cell Biology, VU University, 1081HV Amsterdam, The Netherlands.

[4] Structural Genomics Laboratory, Centro de Investigación Príncipe Felipe. E46012 Valencia, Spain.

Manuscript information: 30 text pages, 2 figures

Running title: Membrane protein integration

Keywords: biogenesis, insertion; membrane protein; translocon; transmembrane segment

= These authors contributed equally to this work

* Corresponding author.

Phone: Int+34-963543796.

Fax: Int+34-963544635.

E-mail: Ismael.Mingarro@uv.es;

Abstract

Most integral membrane proteins are targeted, inserted and assembled in the endoplasmic reticulum (ER) membrane. The sequential and potentially overlapping events for membrane protein integration take place at sites termed translocons, which comprise a specific set of membrane proteins acting in concert with ribosomes and, probably, molecular chaperones to ensure the success of the whole process. In this minireview, we summarize our current understanding of helical membrane protein integration at the ER and highlight specific characteristics that affect the biogenesis of multi-spanning membrane proteins.

Abbreviations:

ER, endoplasmic reticulum; SRP, signal recognition particle; SS, signal sequence; TM, transmembrane; TRAM, translocating chain-associated membrane protein.

Introduction

Helical integral membrane proteins have essential roles in the cell and account for almost one-fourth of all proteins in most organisms [1]. By contrast, our understanding of their biosynthesis and folding lags far behind our understanding of water-soluble proteins. The targeting and insertion of most integral membrane proteins in eukaryotic cells occur co-translationally, where protein synthesis and integration into the endoplasmic reticulum (ER) membrane are coupled. In this case, the targeting of the ribosome-mRNA-nascent chain complex to the membrane depends on the signal recognition particle (SRP) and its interaction with the membrane-bound SRP receptor [2], which is located in close proximity to the translocon. The translocon, a multi-protein complex, facilitates the insertion of integral membrane proteins into the lipid bilayer [3] and the translocation of soluble proteins into the ER lumen [4]. During insertion, nascent membrane proteins have to adopt the correct orientation in the lipid bilayer, undergo covalent modifications (*e.g.* signal sequence cleavage and *N*-linked glycosylation), fold properly, and interact with ER-resident proteins (*e.g.* chaperones), to eventually adopt the their native state. All these series of sequential (and potentially) overlapping events take place in a very peculiar environment, the membrane, where physics significantly differ from the aqueous environment. Therefore, characterizing how membrane proteins integrate into the ER membrane requires detailed knowledge of the constraints imposed by the hydrophobic lipid bilayer as well as its response to accommodate the transmembrane (TM) segments of integral proteins. In this review, we focus on recent advances in

our understanding of the targeting, insertion and folding of mammalian integral membrane proteins.

Targeting to the ER

Co- versus post-translational insertion

Protein targeting to the ER membrane can occur co- or post-translationally depending on the hydrophobicity and location of the signal sequence (SS), which consist of a short span of hydrophobic residues flanked by a positively charged N-terminal and a polar but uncharged C-terminal region ([5], reviewed by [6]). In the co-translational process, targeting of secretory and membrane proteins is mediated by the conserved signal recognition particle (SRP). The eukaryotic SRP, of which the mammalian particle is the best characterized, is composed of a 300 nucleotides 7S RNA and six protein subunits with molecular masses of 9, 14, 19, 54, 68, and 72 kDa (for review see [2, 7]). Among SRP proteins, only SRP54 is highly conserved in all kingdoms of life, being essential for the SRP function [7]. Two domains, the M-domain and the NG-domain, compose SRP54. The M-domain (methionine-rich domain) associates with SRP-RNA and provides the SS binding site while the NG-domain is responsible for GTP binding (G-domain) and the interaction with the ribosome (N-domain). The SRP complex binds to a hydrophobic domain (either a N-terminal SS or a TM segment) in the nascent polypeptide as it emerges from the ribosome [8]. SRP transiently arrests protein synthesis [9] and docks the ribosome-nascent chain-SRP complex to the ER membrane

via the SRP receptor (SR) [10]. SR is a heterodimer formed by the GTPases SRα and SRβ. SRα is structurally and functionally related to SRP54, also containing an NG-domain [11]. Interaction between the SRP and the SR requires GTP binding to both complexes. Subsequently, the ribosome-nascent chain (RNC) is transferred from the SRP to the Sec61 translocon and, GTP hydrolysis triggers SRP-SR dissociation [12]. Structural studies of the RNC-SRP-SR complex reveal that SR interacts with both the ribosome and SRP rendering conformational changes in SRP that favor the RNC transfer to the translocon [13]. Recent studies with prokaryotic homologues disclose an active role of the SRP RNA in coordinating the SRP-SR interactions and the GTP hydrolysis [14, 15]. The SRP disassembly resumes translation and membrane proteins are laterally released by the Sec translocon into the membrane bilayer, while secretory proteins are threaded through the Sec61 machinery. Despite the increasing mechanistic and structural insights on co-translational targeting, we have limited knowledge on how SRP regulates its binding to a diverse set of signal sequences as well as the conformational changes induced by the SR binding that result in transfer of the nascent chain to the translocon [16].

In the post-translational route, proteins are targeted and inserted (or translocated) after translation by cytosolic ribosomes. In yeast, where this pathway is especially prominent, a dedicated complex, termed Sec62/Sec63 complex (also present in mammalian cells), cooperates with the Sec61 translocon in post-translational translocation of soluble (secretory) proteins [17]. In this pathway, cytosolic Hsp40 and Hsp70-type chaperones maintain

polypeptides in a translocation competent state [18], while several luminal chaperones are required to pull the precursor across the membrane [19]. Another subset of proteins is targeted post-translationally to the ER membrane by the TRC40/GET pathway. This subset of proteins are membrane proteins with a C-terminal TM segment, also known as tail-anchored (TA) proteins (recently reviewed by [20]). Although remarkable progress has been made in the identification of targeting factors, the molecular basis underlying TA membrane protein integration remain to be fully clarified. The two post-translational targeting mechanisms appear to be more complex than co-translational biogenesis of membrane proteins. Hence, up to three distinct targeting pathways have been described so far, the SRP-mediated pathway, the ATP-dependent Hsp40/Hsc70-mediated pathway, and the TRC40/GET pathway, also dependent on ATP hydrolysis (for review see [21]).

Translocon structure

The translocon complex is responsible for the insertion of most integral membrane proteins into the lipid bilayer as well as for the translocation of secretory proteins across the ER membrane [4]. The gating capability of this complex in two directions, (that is, across the membrane and laterally into the lipid bilayer), differentiates it from the rest of the cellular channels. In mammalian cells, this proteinaceous complex is composed by the Sec61 α, β and γ subunits plus the translocating chain-associating membrane protein (TRAM) [22]. Since translocon activity can be reproduced by *ab initio*

reconstitution of these four membrane proteins in pure lipids [23], these proteins constitute the core components of the mammalian translocon [3].

Sec61 complex

The eukaryotic Sec61 complex is a heterotrimeric membrane protein complex (Sec61α, Sec61β and Sec61γ), called SecYEG in bacteria and archea. On the one hand, the α- and γ-subunits are highly conserved in all kingdoms and required for survival, both in *E. coli* and *S. cerevisiae*. The β-subunit on the other hand, is not required and does not have significant sequence homology between eukaryotes and eubacteria. The high-resolution structure of the mammalian Sec61 is not yet available. However, we have the homologous structures from *Methanococcus jannaschii* [24], *Thermus thermophilus* [25], *Thermotoga maritima* [26] and *Pyrococcus furiosus* [27], the last two lacking the non-essential β-subunit. The fitting of the crystal structure of SecYEβ from *M. jannaschii* into the cryo–electron microscopy (cryo-EM) density map of an active mammalian Sec61 [28], and of the cryo-EM structure of SecYEG from *E. coli* with the mammalian Sec61 in a resting state [29] indicate a high degree of structural similarity between all Sec complexes.

The α-subunit

Sec61α constitutes the protein-conducting channel of the translocon complex crossing the membrane ten times, with both its N- and C-termini facing the cytosol. Viewed from the top, the protein adopts a square shape that can be divided in two pseudo-symmetric halves, the N-terminal halve containing

TMs 1-5 and the C-terminal comprising TMs 6-10 (red and blue TM segments in Fig 1, respectively). These two parts form an indentation in the centre through which the nascent chain passes and is aligned with the ribosomal exit tunnel [28]. From a lateral view, Sec61α has a rectangular contour and the channel within an hourglass shape [30]. In an inactive state the cytoplasmic entry to the channel has a diameter of approximately 20-25Å [24]. Close to the middle of the membrane the translocation pore reach its narrowest point (5-8Å) constituted by a ring of bulky hydrophobic residues followed by a short helix (TM 2a) that blocks the channel pore (Fig 1). After this "plug", the channel widens again towards the ER lumen. Nevertheless, it has been reported a significant increase in the pore diameter [31], which is probably needed to accommodate the multiple TM segments of multi-spanning nascent chains that may leave the translocon in pairs or groups (see below).

The β-subunit

The β-subunit is the smallest component of the Sec61 complex. It contains a single TM domain located next to TM segments 1 and 4 of Sec61α (Fig. 1A). Although this subunit is not essential either for translocation across ER membrane or insertion of TM segments into the lipid bilayer, it has been described to kinetically facilitate co-translational translocation [32], and to interact with the SR heterodimer probably facilitating recognition of unoccupied translocons by the RNC-SRP-SR complex [33]. The participation of Sec61β in the translocation process is also supported by its direct interaction with the nascent chain and the ribosome [34].

The γ-subunit

Sec61γ has two helices connected by an extended loop (Fig. 1). The first helical region, an amphipathic helix, sits parallel to the cytosolic side of the membrane and contacts with the cytoplasmic side of the Sec61α C-terminal halve. The second helix crosses the membrane diagonally interacting with both N- and C-terminal parts of Sec61α and acts as a clamp that brings both halves of Sec61α together [24].

Translocation and insertion of a nascent chain.

During co-translational insertion/translocation the nascent polypeptide is extruded into the translocon from the ribosome exit tunnel. The precise stoichiometry and structure of the actively engaged translocon-ribosome complex has been a great controversy over the past years. Initial cryo-EM studies indicated that 3-4 copies of the Sec61 complex could interact with the ribosome at the same time [35]. However, biochemical studies and the recent structures available strongly suggest that only one copy of the Sec61/SecY complex is required for translocation [24, 27-29, 36, 37]. Biochemical analysis of Sec61 point mutants [38], and the cryo-EM reconstructions of the ribosome-translocon pair indicate that the loops between TM segments 6-7 and 8-9 of the translocon are involved in this association [28, 39]. In fact, point mutations within those loops of the *E. coli* SecY are known to affect the ribosome-SecY interaction [39]. However, similar changes in loop L6 of the yeast translocon

did not affect binding to ribosome [28]. All in all indicates that, despite small differences, the ribosome-Sec junction is well conserved among species.

Although many details remain unknown, significant insight into the mechanism of membrane insertion has come from structural studies. The process starts with the engagement between the translocon complex and its cytosolic partner (that is, the ribosome in the co-translational pathway). Either this contact or the presence of the SS triggers the widening of the cytosolic side of the channel [25], including the hydrophobic ring that increases from ~5 to ~14Å [27]. In this pre-open state, displacement of TM segments 6, 8 and 9 from their position in the closed configuration would create a lateral "crack" between the two halves of Sec61α (*i.e.*, at the TM segments 2b and 7/8 interface), which would occur only in the cytosolic side of the channel. However, segment 2a retains its location keeping intact the permeability barrier. Once the SS enters into the channel as a loop, its first amino acids interact with the cytosolic residues of TM segment 8. At the same time, the hydrophobic core of the SS contacts TM segments 7 and 2b on both sides of the channel and with the phospholipids through the already open lateral crack [40]. As the elongation of the nascent chain continues two rearrangements occur in Sec61α. First, the plug should be displaced to leave room for the nascent polypeptide, which can now completely expand the channel. Second, the pairs formed by TM segments 2-3 on one side and 7-8 on the other halve move apart from each other (Fig. 1B) creating a lateral gate across the entire channel, which exposes the nascent polypeptide to the core of the membrane [27, 41]. The sequence within the translocon can then

partition into the lipids if it is hydrophobic enough (see TM domain requirements section), as the SS would do, or continue through the translocon into the ER lumen. The structural changes in the α-subunit are accompanied by a dramatic shift (Fig. 1B) in the location of the N-terminal helix form Sec61γ/SecE [27], which releases the clamp over Sec61α. Nevertheless, the opening of the lateral gate is not required to accommodate a translocating peptide within the channel [28]. Therefore, it is possible that the opening of the lateral gate is triggered by the presence of a TM segment inside the translocon, which would adjust its dynamic structure according to the nature of the polypeptide within the channel. During this process, the permeability barrier is kept by the coordinated *in* and *out* movement of the "plug" and the widening/narrowing of the hydrophobic ring, whilst the opening/closing of the lateral gate exposes hydrophobic segments to the lipid bilayer allowing their partition into the membrane.

TRAM

The TRanslocating chain-Associating Membrane protein (TRAM) was identified by crosslinking methods in reconstituted proteoliposomes [22]. Despite being recognized as an essential component for the translocation or insertion into the membrane of several secreted and membrane proteins, its precise function remains unknown. TRAM is an integral membrane protein with 8 TMs and both N- and C-termini facing the cytosol [42]. The role of TRAM in the translocation of secretory proteins is restricted to the insertion of the SS into the membrane [43], where TRAM has been found required for the

insertion of SS with either short hydrophobic sequences or with low overall hydrophobicity. Regarding the insertion of TM segments, TRAM has also been reported to crosslink with a wide variety of TM segments [44-48], some of them containing charged residues [49-51]. These observations together with the fact that TRAM itself contains an unusual high number of charged residues within its TM segments, led to the idea that TRAM could act as a chaperone for the integration of non-optimal TM segments by providing a more favorable context [42].

Translocon-associated proteins

Some other membrane proteins (*i.e.* TRAP, PAT-10, RAMP4 or BAP31) have been reported to interact with the translocon and modulate its function at some stage. However, their presence is not required for either insertion or translocation and thus they are not considered as a part of the translocon core complex.

The TRanslocon Associated Protein (TRAP) is a tetrameric complex (α, β, γ and δ) of integral membrane proteins [52]. It is associated with ribosome-Sec61 complexes with a 1:1 stoichiometry [29]. It has been proposed that TRAP facilitates the initiation of protein translocation [53] although the details of the mechanism remain unknown. PAT-10 was discovered as a translocon-associated protein when looking for Sec61 partners during opsin nascent chain insertion [50]. It is a membrane protein that cross-links with some of the opsin TM segments [54]. This interaction is independent of the presence of *N*-glycosylation sites, the amino acid sequence or the topology of its first TM

segment. Apparently, PAT-10 binding is triggered by the relative location of this TM segment within the opsin nascent chain. RAMP4 was also found tightly associated with the translocon [23]. RAMP4 is a small (66 amino acid) TA membrane protein implicated in promoting correct integration/folding of integral membrane proteins by facilitating subsequent glycosylation [55]. In a translating ribosome-translocon complex, RAMP4 is recruited to the Sec61 complex before the TM segment emerges from the ribosome exit tunnel; hence, it has been postulated that it is the presence of a TM sequence within the ribosome what triggers this recruitment [56]. Another protein that has been reported to interact with the translocon complex is BAP31. This multi-spanning integral membrane protein participates in the identification of misfolded proteins at the ER and their retro-translocation to the cytoplasm. The finding that BAP31 interacts both Sec61β and TRAM [57] suggests a role of the translocon in membrane protein quality control. The increasing amount of interacting partners of the translocon also indicates that different functions of the channel may be performed in association with different cellular components. Indeed, the Sec61 complex might just be the common player in a wide variety of transient complexes each one performing different but related functions.

Transmembrane domain requirements

Hydrophobicity

Individual TM helices follow an ordered insertion pathway, in which they pass from the tunnel in the large ribosomal subunit into the Sec61 translocon

channel and then exit the channel laterally into the surrounding lipids [30, 58]. Generally, the hydrophobicity of the TM sequence drives integration into the membrane. However, the insertion efficiency of TM segments by the translocon depends on amino acid composition, the positions of residues within the segment, TM segment orientation, and the helix length [59-62], suggesting that membrane insertion is fundamentally a fine-tuned thermodynamic partitioning process. Several TM segments from multi-spanning membrane proteins contain charged amino acids that are nevertheless tolerated in the membrane [63, 64]. Computational modeling suggests that integration of TM sequences with a central ionized residue might be assisted by helix–helix interactions within the membrane more than the stabilization of this ionized group by the translocon [65]. *In vivo* and *in vitro* studies suggest that the translocon may act as a facilitator in the insertion/selection process [59, 60, 66], where protein–lipid interactions "decide" the successful integration of the TM segment into the membrane through favorable acyl chain solvation [67] also affected by lipid composition [68]. Indeed, recent work in yeast showed that mutations in the hydrophobic constriction ring of Sec61p influence translocation efficiency modifying the hydrophobicity threshold for membrane insertion [69]. Such a mechanism based on lipid-mediated partitioning would accommodate the diversity of sequences that pass through the translocon in route to the membrane. Nevertheless, it has previously been suggested that the translocon complex can act as a chaperone during the integration of non-optimal TM segments. Indeed, a recent observation that ATP-depletion can halt TM segment release

from the translocon into the bilayer strongly supports this chaperone function [70], which supplement the thermodynamic partitioning process.

Amino acid preferences

A recent annotation on amino acid composition of α-helical TM segments showed that there is considerable information in sequence that relates to the intricate contacts between TM segments [71]. Indeed, there is a biased amino acid preferences depending on whether the residue is exposed to the lipid bilayer or to a soluble environment (Fig. 2). Using all annotation in the MPTopo database [72], we selected amino acids from TM segments and compared their occurrence to amino acids in non-TM segments. In total there were 206 proteins with known three-dimensional structure and topology, which had 1,244 TM segments. The total number of amino acids in TM segments was 25,281 compared to a total of 63,107 amino acids in non-TM regions. As previously reported [73] the hydrophobic residues Leu and Ala make up the bulk of the amino acids in the TM segments accounting for one fourth (24,5%) of all amino acids that are inserted through the translocon, but these two residues are also common in the non-TM regions (16,2%). This effect is even more evident for Gly since its prevalence is almost equal in TM or non-TM regions (Fig. 2). Interestingly, charged residues together with Pro are under-represented in TM domains relatively to non-TM regions. This feature is probably meaningful both in terms of hydrophobicity and helicity.

Helical conformation of transmembrane segments

Formation of a α-helix is critical for the membrane insertion of a TM segment. Even the most hydrophobic polypeptides could not insert into lipid bilayers without concomitant secondary structure formation [74]. One of the most intriguing challenges that membrane proteins have to face is desolvation and partitioning of the polar peptide bond from water into the membrane, which is as unfavorable as that of a charged side chain [75]. However, formation of intramolecular hydrogen bonds (*i.e.*, adoption of secondary structure) can compensate the loss of hydrogen bonds between the polypeptide backbone and water molecules (reviewed in [76]). Where does a predestined TM segment adopt its α-helical conformation? According to the two-stage model (see below), TM segments fold during insertion into the membrane and, in case of multi-spanning membrane proteins, before helix association [77]. However, some TM α-helices have been shown to fold already in the ribosomal tunnel [78-81] even before reaching the translocon or inserting into the lipid bilayer, which suggests that the folding inside the ribosome may regulate the fate of the nascent polypeptide.

Integration mechanism in multi-spanning membrane proteins

During biogenesis of multi-spanning membrane proteins, several TM segments in a single polypeptide need to be integrated by the Sec61□translocon. Unfortunately, our knowledge of the molecular mechanism underlying this process is still very limited. During translation and once the SS or a TM segment has reached the translocon, this first hydrophobic segment has to be relocated to accommodate the following TM

segment within the translocon pore. Whether at this point multiple TM segments partition into the membrane sequentially (*i.e.*, each TM segment exit the translocon individually [49]), or several TM segments can accumulate inside or in the proximities of the translocon and be released into the bilayer in pairs or groups [44, 50, 82, 83], is thought to be protein dependant. Recent structural data have shown that in the pre-open state the hydrophobic ring is widened to ~14Å in the direction of the lateral exit site [27], which is enough for accommodating more than one helix, especially because these dimensions could be further increased in a full open state [31]. It is also known that hydrophobic TM segments leave the translocon sequentially from the N-terminus to the C-terminus [82] and less hydrophobic segments interact with other TM segments at early stages of the membrane integration [46, 54, 84, 85]. More hydrophilic TM segments are forced by downstream hydrophobic sequence to adopt a TM disposition [86, 87]. However, whether these hydrophilic helices are spontaneously inserted or assisted by the Sec61 translocon to insert together with their partner helices is still unknown. Nevertheless, it has been suggested that inter-helical interactions are required to neutralize polar groups in TM sequences [76, 88]. Indeed, recent comparison of helix-helix interactions in available membrane protein structures reveal that they constitute one of the most distinctive characteristics of multi-spanning membrane proteins with more than 4 TM segments [89]. These helix-helix interactions might be coordinated *in vivo* by the translocon or its associated proteins. For example, TRAM (see above) plays a role assisting the integration of hydrophobic sequences containing charged

residues [43, 51]. Therefore, unraveling the functions of translocon-associated proteins will provide new insights into the integration mechanism of non-canonical TM segments.

Topology

During integration, nascent membrane proteins have to adopt the correct topology (*i.e.*, it has to define the number of TM segments and their orientation with respect to the plane of the lipid bilayer [90]), which is likely influenced by the translocon. However, whether a TM segment adopts a N-teminal cytosolic or reverse orientation depends on several factors. First, it has been observed that the folding state of an extra-membrane domain preceding a TM segment precludes its translocation and consequently forces the TM segment towards an N-terminal cytoplasmic orientation [91]. Second, the hydrophobicity of the TM sequence influences membrane orientation. For example, highly hydrophobic sequences promote N-terminal translocation despite the presence of moderate hydrophobic TM segments favor the opposite orientation [92]. Third, and most important, it has been long known that the distribution of charged residues between the flanking regions of a TM segment is a major determinant of topology in membrane proteins [93, 94]. The so-called 'positive-inside rule' was first observed for prokaryotic proteins, where bacteria maintain a net negative-inside electrical potential across the membrane, and a cytoplasmic bilayer leaflet enriched in negatively charged lipids also promotes charge bias. A similar skewed distribution was also identified later in eukaryotes [95], where the balance between positive and

negative charges drives protein topology. Indeed, changing the flanking charges by site-directed mutagenesis can reverse the topology of a TM segment (reviewed in [96]). Moreover, it has been recently demonstrated that certain residues of the translocon also contribute to the positive-inside orientation of signal sequences [97, 98]. Therefore, the amino acid sequence appears to be the primary determinant of final topology, which should be initially interpreted by the translocon. Nevertheless, it has also been reported that membrane lipid composition also influences the final topological orientation of membrane proteins (reviewed in [99]). In summary, both the amino acid sequence of a membrane protein and the collective determinants in the bilayer membrane influence protein topology.

Multi-spanning membrane proteins generally adopt their native orientation depending on the insertion of the SS or the first TM segment, which determines the alternant insertion of the rest of the protein. Nevertheless, drastic changes in loop regions that favor inverted orientations have only local effects [100]. Furthermore, its has been recently shown that the topology of a full length protein can be changed by simply adding a positively charged residue irrespectively of the region of the protein where the mutation was placed, including the C-terminal end of the protein [101]. Unfortunately, the molecular mechanisms by which downstream determinants contribute to the topology is yet unknown [102]. Therefore, experimental evidences are now challenging the classic static view for attaining membrane protein topology. For example, some proteins may adopt multiple topologies depending on the cellular localization or environment

[103] while others like viral membrane proteins have a strong preference for a specific topology [104].

Hydrophobic matching

The effect of the so-called hydrophobic matching on the assembly and orientation of TM segments has been widely studied [105]. A 'mismatch' happens when the hydrophobic thickness of the membrane do not match the length of the hydrophobic region of a TM segment [106]. There has been two types of hydrophobic mismatch described: (i) positive, when the membrane is not thick enough for a TM segment, and (ii) negative, when the length of the hydrophobic section of a TM segment is too short to span the hydrophobic core of the lipid bilayer. In both scenarios, either the membrane or the polypeptide will adapt to minimize the exposure of hydrophobic residues to the aqueous media (positive mismatch) or the extrusion of polar amino acids within the hydrophobic core of the membrane (negative mismatch) [107]. Both re-arrangements are known to be important for determining the final assembly of a membrane protein as proved by fluorescence [108-110] and chimeric overexpression of dimerizing TM segments in membrane-mimetic environments [111, 112]. The ability of the Sec61 translocon to handle negative mismatch has recently been studied [62]. In this work, it has been demonstrated that polyleucine segments as short as ~10 residues long integrate efficiently into the ER membrane. Finally, hydrophobic matching may reflect an evolutionary strategy to regulate the activity of membrane

proteins by allowing the adaptation of TM segment lengths to bilayer thickness in different cellular membranes [113].

Folding and assembly of multi-spanning membrane proteins

Forces behind the folding of membrane proteins

Next, we briefly introduce the molecular interactions driving protein folding within membranes. For a recent complete review see [74, 75]. While hydrophobic collapse is a major driving force in the folding of soluble proteins, its role in membrane proteins is mostly limited to the formation of secondary structures across the lipid bilayer. Similarly, salt bridges and aromatic interactions do not have a weighty contribution in membrane protein folding. Conversely, inter-helical hydrogen bonding [114, 115] and specially van der Waals forces have been identified as major promoters of membrane protein folding [116, 117]. Therefore, the restrictions imposed by the lipid bilayer allow for an effective folding of TM segments of integral membrane proteins despite the low contribution of hydrophobic forces and the reduced effect of salt bridges and aromatic interactions [118].

Folding and assembly of membrane proteins: The two-stage model.

The folding and assembly of helical membrane proteins was schematized more than two decades ago as a two-stage process [77]. First, each TM helix is formed and independently inserted into the lipid bilayer. Second, these helices interact with each other to establish the final structure of the protein.

Although this simplified view has been refined in the past years still constitutes a valid conceptual approach.

In vivo, the insertion into the ER membrane occurs co-translationally via the translocon complex. In this scenario a TM segment does not insert into the membrane spontaneously, instead the translocon facilitates its partition from the aqueous environment within the translocon pore into the lipid bilayer. After insertion, or for some proteins during insertion, the TM helices interact with each other to form higher order structures. These interactions create a microenvironment that permits further changes in the protein structure, such as insertion into the membrane of re-entrant loops or short polypeptides, membrane packing of non-α-helical segments and binding of prosthetics groups [119].

Finally, the influence of the specific lipid environment during the assembly of TM segments should also be taken into account. The lipid and protein components of biological membranes have co-evolved allowing membrane proteins to assemble and function in the heterogenic environment provided by the diverse lipid bilayers in a cell. Not only membrane thickness (see Hydrophobic matching section) but also membrane lateral pressure [120], charge density [121], and even unique lipid-protein interactions [122] have been identified as structural determinants of membrane proteins. Furthermore, very recent cryo-EM studies using RNC complexes bound to SecY reconstituted in nano-discs revealed an interaction of the ribosome with lipids, leading to a disorder in the lipid microenvironment adjacent to the translocon, which may contribute to favor membrane insertion of TM

segments [123]. All in all, the final structure of a multi-spanning membrane protein will not solely be defined by protein-protein and lipid-protein interactions but also by the folding of its soluble domains. Thus, the aqueous environment on both sides of the membrane imposes restrictions to the folding of the extra-membranous regions, and by extension to the overall protein structure.

Concluding Remarks

Membrane protein integration appears to be orchestrated by multiple determinants and factors that in unlimited combinations give raise to native protein structures. During protein targeting, TM segment insertion and assembly into the membrane several inter-connected processes occur simultaneously. Structural studies of the translocon together with *in vitro* quantitative thermodynamic analyses and biophysical dissection of TM interactions have resulted in significant advance of our understanding of membrane protein integration into the lipid bilayer. Our current knowledge coupled with bioinformatics analysis [124] is opening now opportunities for *de novo* membrane protein structure prediction and design.

Acknowledgments

The authors acknowledge financial support from the Spanish Ministry of Science and Innovation (BFU2009-08401/BMC to I.M. and BFU2010-

19310/BMC to M.A.M-R), and from the Generalitat Valenciana (PROMETEO/2010/005 and ACOMP/2011/025 to I.M., and ACOMP/2011/048 to M.A.M-R.).

Figure legends

Fig. 1. Translocon Structure. Top view of the translocon structure. A) Closed structure of the translocon from *M. jannaschii* (PDB ID code 1RHZ) [20]. B) Partially open structure of the translocon from *P. furiosus* (PDB ID code 3MP7) [23]. In both panels all TM segments of Sec61α are colored (red and blue for each half, see text) except for the β and γ subunits, which are shown in grey. All TM segments are numbered for easy comparison between the open and closed structures. Dotted arrows in panel B indicate helix displacements required for the widening of the channel and opening of the lateral gate. A solid arrow shows the lateral gate exit pathway of a TM segment from the interior of the channel into the membrane.

Fig. 2. Amino acid preferences in TM segments compared to loop regions (non-TM) in membrane protein structures. Top two rows show the percentage of occurrence of all amino acid types in TM segments and non-TM segments in membrane proteins of known structure. Lower plot shows the log odds ratio of the occurrence. Briefly, a log odd ratio the is the $\log_{10}$ ratio of the odds of an amino acid occurring in TM segments to the odds of it occurring in a non-TM segment. Positive log odds indicate over-occurrence of the amino acid type in TM segments. Negative log odds indicate under-representation of the amino acid type in TM segments. Amino acids are colored according to an arbitrary division of their log odds (*i.e.*, green color for log odds > 0.3; orange color for 0.3 • log odds • -0.3; and red color for log odds < -0.3).

References

1. Wallin E & von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-1038.
2. Egea PF, Stroud RM & Walter P (2005) Targeting proteins to membranes: structure of the signal recognition particle. *Current Opinion in Structural Biology* **15**, 213-220.
3. Johnson AE & van Waes MA (1999) The translocon: A dynamic gateway at the ER membrane. *Annual Review of Cell and Developmental Biology* **15**, 799-842.
4. Rapoport TA (2008) Protein transport across the endoplasmic reticulum membrane. *FEBS J* **275**, 4471-4478.
5. Ng DTW, Brown JD & Walter P (1996) Signal sequences specify the targeting route to the endoplasmic reticulum. *J Cell Biol* **134**, 269-278.
6. Cross BC, Sinning I, Luirink J & High S (2009) Delivering proteins for export from the cytosol. *Nat Rev Mol Cell Biol* **10**, 255-264.
7. Pool MR (2005) Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). *Mol Membr Biol* **22**, 3-15.
8. Halic M, Blau M, Becker T, Mielke T, Pool MR, Wild K, Sinning I & Beckmann R (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**, 507-511.
9. Lakkaraju AKK, Mary C, Scherrer A, Johnson AE & Strub K (2008) SRP Keeps Polypeptides Translocation-Competent by Slowing Translation to Match Limiting ER-Targeting Sites. *Cell* **133**, 440-451.
10. Gilmore R, Blobel G & Walter P (1982) Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J Cell Biol* **95**, 463-469.
11. Montoya G, Svensson C, Luirink J & Sinning I (1997) Crystal structure of the NG domain from the signal recognition particle receptor FtsY. *Nature* **385**, 365-368.
12. Song W, Raden D, Mandon EC & Gilmore R (2000) Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell* **100**, 333-343.
13. Halic M, Gartmann M, Schlenker O, Mielke T, Pool MR, Sinning I & Beckmann R (2006) Signal Recognition Particle Receptor Exposes the Ribosomal Translocon Binding Site. *Science* **312**, 745-747, doi: 10.1126/science.1124864.
14. Janda CY, Li J, Oubridge C, Hernandez H, Robinson CV & Nagai K (2010) Recognition of a signal peptide by the signal recognition particle. *Nature* **465**, 507-510.
15. Ataide SF, Schmitz N, Shen K, Ke A, Shan S-o, Doudna JA & Ban N (2011) The Crystal Structure of the Signal Recognition Particle in Complex with Its Receptor. *Science* **331**, 881-886.
16. Hainzl T, Huang S, Merilainen G, Brannstrom K & Sauer-Eriksson AE (2011) Structural basis of signal-sequence recognition by the signal recognition particle. *Nat Struct Mol Biol* **18**, 389-391.
17. 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.
18. Becker J, Walter W, Yan W & Craig EA (1996) Functional interaction of cytosolic hsp70 and a DnaJ- related protein, Ydj1p, in protein translocation in vivo. *Mol Cell Biol* **16**, 4378-4386.

19. Wilkinson BM, Regnacq M & Stirling CJ (1997) Protein translocation across the membrane of the endoplasmic reticulum. *J Membr Biol* **155**, 189-197.
20. Borgese N & Fasana E (2011) Targeting pathways of C-tail-anchored proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 937-946.
21. Rabu C, Schmid V, Schwappach B & High S (2009) Biogenesis of tail-anchored proteins: the beginning for the end? *J Cell Sci* **122**, 3605-3612.
22. Görlich D, Hartmann E, Prehn S & Rapoport TA (1992) A Protein of the Endoplasmic Reticulum Involved Early in Polypeptide Translocation. *Nature* **357**, 47-52.
23. Görlich D & Rapoport TA (1993) Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615-630.
24. Van den Berg B, Clemons WM, Jr., Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.
25. Tsukazaki T, Mori H, Fukai S, Ishitani R, Mori T, Dohmae N, Perederina A, Sugita Y, Vassylyev DG, Ito K, et al. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455**, 988-991.
26. Zimmer J, Nam Y & Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936-943.
27. Egea PF & Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci U S A* **107**, 17182-17187.
28. Becker T, Bhushan S, Jarasch A, Armache JP, Funes S, Jossinet F, Gumbart J, Mielke T, Berninghausen O, Schulten K, et al. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369-1373.
29. Menetret JF, Hegde RS, Aguiar M, Gygi SP, Park E, Rapoport TA & Akey CW (2008) Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16**, 1126-1137.
30. Rapoport TA, Goder V, Heinrich SU & Matlack KE (2004) Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol* **14**, 568-575.
31. Hamman BD, Chen J-C, Johnson EE & Johnson AE (1997) The aqueous pore through the translocon has a diameter of 40-60 Å during cotrsnaltional protein translocation at the ER membrane. *Cell* **89**, 535-544.
32. Kalies KU, Rapoport TA & Hartmann E (1998) The beta subunit of the Sec61 complex facilitates cotranslational protein transport and interacts with the signal peptidase during translocation. *J Cell Biol* **141**, 887-894.
33. Jiang Y, Cheng Z, Mandon EC & Gilmore R (2008) An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol* **180**, 1149-1161, doi: jcb.200707196 [pii] 10.1083/jcb.200707196.
34. Levy R, Wiedmann M & Kreibich G (2001) In vitro binding of ribosomes to the beta subunit of the Sec61p protein translocation complex. *J Biol Chem* **276**, 2340-2346.
35. Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G & Frank J (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex [see comments]. *Science* **278**, 2123-2126.
36. Yahr TL & Wickner WT (2000) Evaluating the oligomeric state of SecYEG in preprotein translocase. *EMBO J* **19**, 4393-4401.

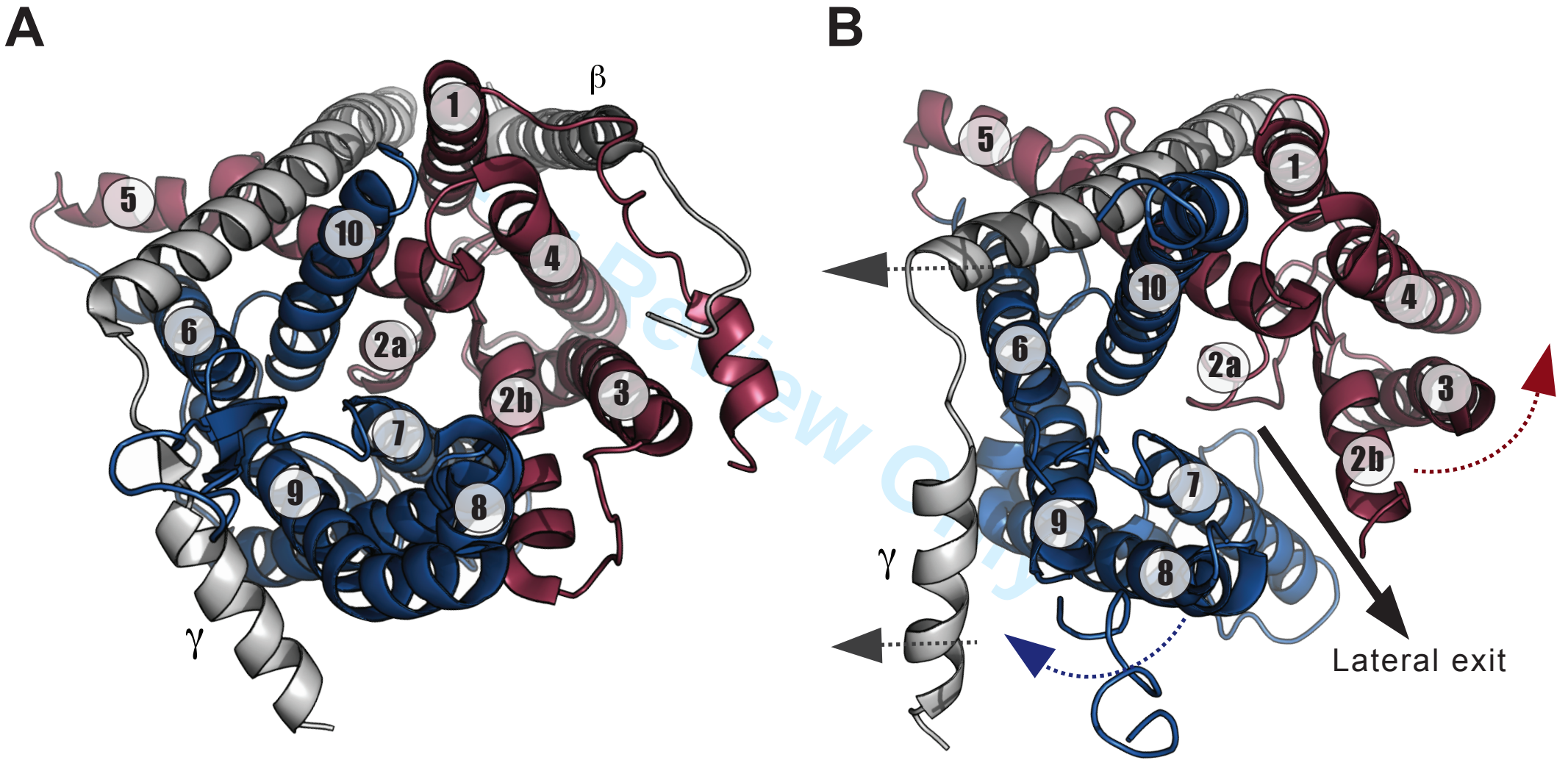
37. Cannon KS, Or E, Clemons WM, Jr., Shibata Y & Rapoport TA (2005) Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J Cell Biol* **169**, 219-225.
38. Cheng Z, Jiang Y, Mandon EC & Gilmore R (2005) Identification of cytoplasmic residues of Sec61p involved in ribosome binding and cotranslational translocation. *J Cell Biol* **168**, 67-77.
39. Menetret JF, Schaletzky J, Clemons WM, Jr., Osborne AR, Skanland SS, Denison C, Gygi SP, Kirkpatrick DS, Park E, Ludtke SJ, et al. (2007) Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol Cell* **28**, 1083-1092.
40. Plath K, Mothes W, Wilkinson BM, Stirling CJ & Rapoport TA (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94**, 795-807.
41. Martoglio B, Hofmann MW, Brunner J & Dobberstein B (1995) The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell* **81**, 207-214.
42. Tamborero S, Vilar M, Martinez-Gil L, Johnson AE & Mingarro I (2011) Membrane Insertion and Topology of the Translocating Chain-Associating Membrane Protein (TRAM). *J Mol Biol* **406**, 571-582.
43. Voigt S, Jungnickel B, Hartmann E & Rapoport TA (1996) Signal sequence-dependent function of the TRAM protein during early pphases of protein transport across the endoplasmic reticulum membrane. *J Cell Biol* **134**, 25-35.
44. McCormick PJ, Miao Y, Shao Y, Lin J & Johnson AE (2003) Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. *Mol Cell* **12**, 329-341.
45. Do H, Falcone D, Lin J, Andrews DW & Johnson AE (1996) The cotransational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* **85**, 369-378.
46. Heinrich SU & Rapoport TA (2003) Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane. *Embo J* **22**, 3654-3663.
47. Sauri A, McCormick PJ, Johnson AE & Mingarro I (2007) Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. *J Mol Biol* **366**, 366-374.
48. Martinez-Gil L, Johnson AE & Mingarro I (2010) Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *J Virol* **84**, 5520-5527.
49. Heinrich SU, Mothes W, Brunner J & Rapoport TA (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**, 233-244.
50. Meacock SL, Lecomte FJ, Crawshaw SG & High S (2002) Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol Biol Cell* **13**, 4114-4129.
51. Cross BC & High S (2009) Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J Cell Sci* **122**, 1768-1777.
52. Hartmann E, Gorlich D, Kostka S, Otto A, Kraft R, Knespel S, Burger E, Rapoport TA & Prehn S (1993) A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur J Biochem* **214**, 375-381.
53. Fons RD, Bogert BA & Hegde RS (2003) Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J Cell Biol* **160**, 529-539.

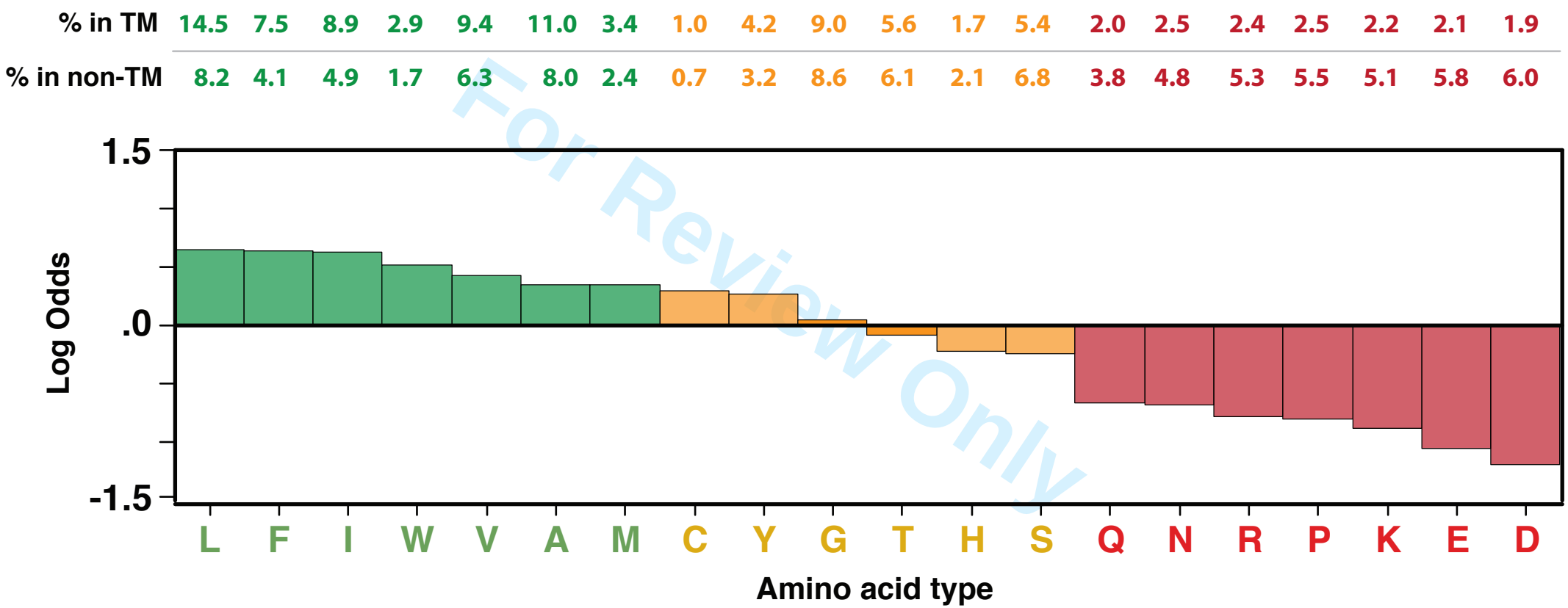
54. Ismail N, Crawshaw SG, Cross BC, Haagsma AC & High S (2008) Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. *Biochem J* **411**, 495-506.
55. Yamaguchi A, Hori O, Stern DM, Hartmann E, Ogawa S & Tohyama M (1999) Stress-Associated Endoplasmic Reticulum Protein 1 (Serp1)/Ribosome-Associated Membrane Protein 4 (Ramp4) Stabilizes Membrane Proteins during Stress and Facilitates Subsequent Glycosylation. *The Journal of Cell Biology* **147**, 1195-1204.
56. Pool MR (2009) A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J Cell Biol* **185**, 889-902.
57. Wang B, Heath-Engel H, Zhang D, Nguyen N, Thomas DY, Hanrahan JW & Shore GC (2008) BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTRDeltaF508 via the derlin-1 complex. *Cell* **133**, 1080-1092.
58. Alder NN & Johnson AE (2004) Cotranslational membrane protein biogenesis at the endoplasmic reticulum. *J Biol Chem* **279**, 22787-22790.
59. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH & von Heijne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-381.
60. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH & von Heijne G (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030.
61. Lundin C, Kim H, Nilsson I, White SH & von Heijne G (2008) Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for Nin/Cout and Nout/Cin transmembrane helices. *Proceedings of the National Academy of Sciences* **105**, 15702-15707.
62. Jaud S, Fernandez-Vidal M, Nilsson I, Meindl-Beinker NM, Hubner NC, Tobias DJ, von Heijne G & White SH (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proc Natl Acad Sci U S A* **106**, 11588-11593.
63. Martinez-Gil L, Perez-Gil J & Mingarro I (2008) The surfactant peptide KL4 sequence is inserted with a transmembrane orientation into the endoplasmic reticulum membrane. *Biophys J* **95**, L36-38.
64. White SH & von Heijne G (2008) How translocons select transmembrane helices. *Annu Rev Biophys* **37**, 23-42.
65. Rychkova A, Vicatos S & Warshel A (2010) On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. *Proceedings of the National Academy of Sciences* **107**, 17598-17603.
66. Mulvihill CM & Deber CM (2010) Evidence that the translocon may function as a hydropathy partitioning filter. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1798**, 1995-1998.
67. Johansson AC & Lindahl E (2009) Titratable amino acid solvation in lipid membranes as a function of protonation state. *J Phys Chem B* **113**, 245-253.
68. Johansson AC & Lindahl E (2009) The role of lipid composition for insertion and stabilization of amino acids in membranes. *J Chem Phys* **130**, 185101.
69. 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.
70. Pitonzo D, Yang Z, Matsumura Y, Johnson AE & Skach WR (2009) Sequence-Specific Retention and Regulated Integration of a Nascent Membrane Protein by the ER Sec61 Translocon. *Mol Biol Cell* **20**, 685-698.

71. Arce J, Sturgis JN & Duneau JP (2009) Dissecting membrane protein architecture: An annotation of structural complexity. *Biopolymers* **91**, 815-829.
72. Jayasinghe S, Hristova K & White SH (2001) MPtopo: A database of membrane protein topology. *Protein Sci* **10**, 455-458.
73. Ulmschneider MB & Sansom MS (2001) Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* **1512**, 1-14.
74. Fiedler S, Broecker J & Keller S (2010) Protein folding in membranes. *Cellular and Molecular Life Sciences* **67**, 1779-1798.
75. White SH & Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**, 319-365.
76. Bowie JU (2011) Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol* **21**, 42-49.
77. Popot JL & Engelman DM (1990) Membrane protein folding and oligomerization - The 2-stage model. *Biochemistry* **29**, 4031-4037.
78. Mingarro I, Nilsson I, Whitley P & von Heijne G (2000) Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol* **1**, 3.
79. Woolhead CA, McCormick PJ & Johnson AE (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725-736.
80. Tu LW & Deutsch C A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation. *Journal of Molecular Biology* **396**, 1346-1360.
81. Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, Berninghausen O, Wilson DN & Beckmann R (2010) alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* **17**, 313-317.
82. Sadlish H, Pitonzo D, Johnson AE & Skach WR (2005) Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nat Struct Mol Biol* **12**, 870-878.
83. Sauri A, Saksena S, Salgado J, Johnson AE & Mingarro I (2005) Double-spanning Plant Viral Movement Protein Integration into the Endoplasmic Reticulum Membrane Is Signal Recognition Particle-dependent, Translocon-mediated, and Concerted. *J Biol Chem* **280**, 25907-25912.
84. Ota K, Sakaguchi M, Hamasaki N & Mihara K (2000) Membrane integration of the second transmembrane segment of band 3 requires a closely apposed preceding signal-anchor sequence. *J Biol Chem* **275**, 29743-29748.
85. Hedin LE, Ojemalm K, Bernsel A, Hennerdal A, Illergard K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* **396**, 221-229.
86. Ota K, Sakaguchi M, von Heijne G, Hamasaki N & Mihara K (1998) Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. *Mol Cell* **2**, 495-503.
87. Kida Y, Morimoto F & Sakaguchi M (2007) Two translocating hydrophilic segments of a nascent chain span the ER membrane during multispanning protein topogenesis. *J Cell Biol* **179**, 1441-1452.
88. Hermansson M & von Heijne G (2003) Inter-helical Hydrogen Bond Formation During Membrane Protein Integration into the ER Membrane. *Journal of Molecular Biology* **334**, 803-809.

89. Fuchs A & Frishman D (2010) Structural comparison and classification of alpha-helical transmembrane domains based on helix interaction patterns. *Proteins: Structure, Function, and Bioinformatics* **78**, 2587-2599, doi: 10.1002/prot.22768.
90. von Heijne G (2006) Membrane-protein topology. *Nat Rev Mol Cell Biol* **7**, 909-918.
91. 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.
92. Goder V & Spiess M (2003) Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* **22**, 3645-3653.
93. von Heijne G (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456-458.
94. 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.
95. Hartmann E, Rapoport TA & Lodish HF (1989) Predicting the orientation of eukaryotic membrane proteins. *Proc Natl Acad Sci USA* **86**, 5786-5790.
96. Goder V & Spiess M (2001) Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett* **504**, 87-93.
97. 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.
98. 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.
99. Dowhan W & Bogdanov M (2009) Lipid-dependent membrane protein topogenesis. *Annu Rev Biochem* **78**, 515-540.
100. 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.
101. Seppala S, Slusky JS, Lloris-Garcera P, Rapp M & von Heijne G (2010) Control of membrane protein topology by a single C-terminal residue. *Science* **328**, 1698-1700.
102. Nilsson I, Witt S, Kiefer H, Mingarro I & von Heijne G (2000) Distant Downstream Sequence Determinants Can Control N-tail Translocation during Protein Insertion into the Endoplasmic Reticulum Membrane. *J Biol Chem* **275**, 6207-6213.
103. Hegde RS, Voigt S & Lingappa VR (1998) Regulation of protein topology by trans-acting factors at the endoplasmic reticulum. *Mol Cell* **2**, 85-91.
104. Sauri A, Tamborero S, Martinez-Gil L, Johnson AE & Mingarro I (2009) Viral membrane protein topology is dictated by multiple determinants in its sequence. *J Mol Biol* **387**, 113-128.
105. Dumas F, Lebrun MC & Tocanne JF (1999) Is the protein/lipid hydrophobic matching principle relevant to membrane organization and functions? *FEBS Lett* **458**, 271-277.
106. Killian JA (1998) Hydrophobic mismatch between proteins and lipids in membranes. *Biochim Biophys Acta* **10**, 401-415.
107. Andersen OS & Koeppe RE, 2nd (2007) Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct* **36**, 107-130, doi: 10.1146/annurev.biophys.36.040306.132643.

108. Sparr E, Ash WL, Nazarov PV, Rijkers DTS, Hemminga MA, Tieleman DP & Killian JA (2005) Self-association of Transmembrane Œ±-Helices in Model Membranes. *Journal of Biological Chemistry* **280**, 39324-39331.
109. Mall S, Broadbridge R, Sharma RP, East JM & Lee AG (2001) Self-association of model transmembrane alpha-helices is modulated by lipid structure. *Biochemistry* **40**, 12379-12386.
110. Ren J, Lew S, Wang J & London E (1999) Control of the Transmembrane Orientation and Interhelical Interactions within Membranes by Hydrophobic Helix Length. *Biochemistry* **38**, 5905-5912.
111. Orzaez M, Perez-Paya E & Mingarro I (2000) Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. *Protein Sci* **9**, 1246-1253.
112. Orzaez M, Lukovic D, Abad C, Perez-Paya E & Mingarro I (2005) Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. *FEBS Lett* **579**, 1633-1638.
113. Sharpe HJ, Stevens TJ & Munro S (2010) A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158-169.
114. Joh NH, Min A, Faham S, Whitelegge JP, Yang D, Woods VL & Bowie JU (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266-1270.
115. Popot JL & Engelman DM (2000) Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
116. MacKenzie KR, Prestegard JH & Engelman DM (1997) A transmembrane helix dimer: Structure and implications. *Science* **276**, 131-133.
117. Eilers M, Shekar SC, Shieh T, Smith SO & Fleming PJ (2000) Internal packing of helical membrane proteins. *Proc Natl Acad Sci U S A* **97**, 5796-5801.
118. Bowie JU (2005) Solving the membrane protein folding problem. *Nature* **438**, 581-589.
119. Engelman DM, Chen Y, Chin CN, Curran AR, Dixon AM, Dupuy AD, Lee AS, Lehnert U, Matthews EE, Reshetnyak YK, et al. (2003) Membrane protein folding: beyond the two stage model. *FEBS Lett* **555**, 122-125.
120. van den Brink-van der Laan E, Chupin V, Killian JA & de Kruijff B (2004) Small Alcohols Destabilize the KcsA Tetramer via Their Effect on the Membrane Lateral Pressure. *Biochemistry* **43**, 5937-5942.
121. Bogdanov M & Dowhan W (1999) Lipid-assisted Protein Folding. *Journal of Biological Chemistry* **274**, 36827-36830, doi: 10.1074/jbc.274.52.36827.
122. Lee AG (2003) Lipid-protein interactions in biological membranes: a structural perspective. *Biochim Biophys Acta* **1612**, 1-40.
123. Frauenfeld J, Gumbart J, Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, Schulten K, et al. (2011) Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat Struct Mol Biol* **18**, 614-621.
124. Elofsson A & von Heijne G (2007) Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76**, 125-140.


A
B
β
γ
1
2a
2b
3
4
5
6
7
8
9
10
Lateral exit



13 May 2011

Membrane protein integration into the ER

Luis Martínez-Gil [1,2=], Ana Saurí [3=], Marc A. Marti-Renom [4] and Ismael Mingarro [1] *

[1] Departament de Bioquímica i Biologia Molecular, Universitat de València. E46100 Burjassot, Spain.

[2] Department of Microbiology, Mount Sinai School of Medicine, One Gustave Levy Place, New York, NY 10029-6574, USA.

[3] Department of Molecular Microbiology, Institute of Molecular Cell Biology, VU University, 1081HV Amsterdam, The Netherlands.

[4] Structural Genomics Laboratory, Centro de Investigación Príncipe Felipe. E46012 Valencia, Spain.

Manuscript information: 30 text pages, 2 figures

Running title: Membrane protein integration

Keywords: biogenesis, insertion; membrane protein; translocon; transmembrane segment

= These authors contributed equally to this work

* Corresponding author.

Phone: Int+34-963543796.

Fax: Int+34-963544635.

E-mail: Ismael.Mingarro@uv.es;

Abstract

Most integral membrane proteins are targeted, inserted and assembled in the endoplasmic reticulum (ER) membrane. The sequential and potentially overlapping events for membrane protein integration take place at sites termed translocons, which comprise a specific set of membrane proteins acting in concert with ribosomes and, probably, molecular chaperones to ensure the success of the whole process. In this minireview, we summarize our current understanding of helical membrane protein integration at the ER and highlight specific characteristics that affect the biogenesis of multi-spanning membrane proteins.

Abbreviations:

ER, endoplasmic reticulum; SRP, signal recognition particle; SS, signal sequence; TM, transmembrane; TRAM, translocating chain-associated membrane protein.

Introduction

Helical integral membrane proteins have essential roles in the cell and account for almost one-fourth of all proteins in most organisms [1]. By contrast, our understanding of their biosynthesis and folding lags far behind our understanding of water-soluble proteins. The targeting and insertion of most integral membrane proteins in eukaryotic cells occur co-translationally, where protein synthesis and integration into the endoplasmic reticulum (ER) membrane are coupled. In this case, the targeting of the ribosome-mRNA-nascent chain complex to the membrane depends on the signal recognition particle (SRP) and its interaction with the membrane-bound SRP receptor [2], which is located in close proximity to the translocon. The translocon, a multi-protein complex, facilitates the insertion of integral membrane proteins into the lipid bilayer [3] and the translocation of soluble proteins into the ER lumen [4]. During insertion, nascent membrane proteins have to adopt the correct orientation in the lipid bilayer, undergo covalent modifications (*e.g.* signal sequence cleavage and *N*-linked glycosylation), fold properly, and interact with ER-resident proteins (*e.g.* chaperones), to eventually adopt the their native state. All these series of sequential (and potentially) overlapping events take place in a very peculiar environment, the membrane, where physics significantly differ from the aqueous environment. Therefore, characterizing how membrane proteins integrate into the ER membrane requires detailed knowledge of the constraints imposed by the hydrophobic lipid bilayer as well as its response to accommodate the transmembrane (TM) segments of integral proteins. In this review, we focus on recent advances in

our understanding of the targeting, insertion and folding of mammalian integral membrane proteins.

Targeting to the ER

Co- versus post-translational insertion

Protein targeting to the ER membrane can occur co- or post-translationally depending on the hydrophobicity and location of the signal sequence (SS), which consist of a short span of hydrophobic residues flanked by a positively charged N-terminal and a polar but uncharged C-terminal region ([5], reviewed by [6]). In the co-translational process, targeting of secretory and membrane proteins is mediated by the conserved signal recognition particle (SRP). The eukaryotic SRP, of which the mammalian particle is the best characterized, is composed of a 300 nucleotides 7S RNA and six protein subunits with molecular masses of 9, 14, 19, 54, 68, and 72 kDa (for review see [2, 7]). Among SRP proteins, only SRP54 is highly conserved in all kingdoms of life, being essential for the SRP function [7]. Two domains, the M-domain and the NG-domain, compose SRP54. The M-domain (methionine-rich domain) associates with SRP-RNA and provides the SS binding site while the NG-domain is responsible for GTP binding (G-domain) and the interaction with the ribosome (N-domain). The SRP complex binds to a hydrophobic domain (either a N-terminal SS or a TM segment) in the nascent polypeptide as it emerges from the ribosome [8]. SRP transiently arrests protein synthesis [9] and docks the ribosome-nascent chain-SRP complex to the ER membrane

via the SRP receptor (SR) [10]. SR is a heterodimer formed by the GTPases SRα and SRβ. SRα is structurally and functionally related to SRP54, also containing an NG-domain [11]. Interaction between the SRP and the SR requires GTP binding to both complexes. Subsequently, the ribosome-nascent chain (RNC) is transferred from the SRP to the Sec61 translocon and, GTP hydrolysis triggers SRP-SR dissociation [12]. Structural studies of the RNC-SRP-SR complex reveal that SR interacts with both the ribosome and SRP rendering conformational changes in SRP that favor the RNC transfer to the translocon [13]. Recent studies with prokaryotic homologues disclose an active role of the SRP RNA in coordinating the SRP-SR interactions and the GTP hydrolysis [14, 15]. The SRP disassembly resumes translation and membrane proteins are laterally released by the Sec translocon into the membrane bilayer, while secretory proteins are threaded through the Sec61 machinery. Despite the increasing mechanistic and structural insights on co-translational targeting, we have limited knowledge on how SRP regulates its binding to a diverse set of signal sequences as well as the conformational changes induced by the SR binding that result in transfer of the nascent chain to the translocon [16].

In the post-translational route, proteins are targeted and inserted (or translocated) after translation by cytosolic ribosomes. In yeast, where this pathway is especially prominent, a dedicated complex, termed Sec62/Sec63 complex (also present in mammalian cells), cooperates with the Sec61 translocon in post-translational translocation of soluble (secretory) proteins [17]. In this pathway, cytosolic Hsp40 and Hsp70-type chaperones maintain

polypeptides in a translocation competent state [18], while several luminal chaperones are required to pull the precursor across the membrane [19]. Another subset of proteins is targeted post-translationally to the ER membrane by the TRC40/GET pathway. This subset of proteins are membrane proteins with a C-terminal TM segment, also known as tail-anchored (TA) proteins (recently reviewed by [20]). Although remarkable progress has been made in the identification of targeting factors, the molecular basis underlying TA membrane protein integration remain to be fully clarified. The two post-translational targeting mechanisms appear to be more complex than co-translational biogenesis of membrane proteins. Hence, up to three distinct targeting pathways have been described so far, the SRP-mediated pathway, the ATP-dependent Hsp40/Hsc70-mediated pathway, and the TRC40/GET pathway, also dependent on ATP hydrolysis (for review see [21]).

Translocon structure

The translocon complex is responsible for the insertion of most integral membrane proteins into the lipid bilayer as well as for the translocation of secretory proteins across the ER membrane [4]. The gating capability of this complex in two directions, (that is, across the membrane and laterally into the lipid bilayer), differentiates it from the rest of the cellular channels. In mammalian cells, this proteinaceous complex is composed by the Sec61 α, β and γ subunits plus the translocating chain-associating membrane protein (TRAM) [22]. Since translocon activity can be reproduced by *ab initio*

reconstitution of these four membrane proteins in pure lipids [23], these proteins constitute the core components of the mammalian translocon [3].

Sec61 complex

The eukaryotic Sec61 complex is a heterotrimeric membrane protein complex (Sec61α, Sec61β and Sec61γ), called SecYEG in bacteria and archea. On the one hand, the α- and γ-subunits are highly conserved in all kingdoms and required for survival, both in *E. coli* and *S. cerevisiae*. The β-subunit on the other hand, is not required and does not have significant sequence homology between eukaryotes and eubacteria. The high-resolution structure of the mammalian Sec61 is not yet available. However, we have the homologous structures from *Methanococcus jannaschii* [24], *Thermus thermophilus* [25], *Thermotoga maritima* [26] and *Pyrococcus furiosus* [27], the last two lacking the non-essential β-subunit. The fitting of the crystal structure of SecYEβ from *M. jannaschii* into the cryo–electron microscopy (cryo-EM) density map of an active mammalian Sec61 [28], and of the cryo-EM structure of SecYEG from *E. coli* with the mammalian Sec61 in a resting state [29] indicate a high degree of structural similarity between all Sec complexes.

The α-subunit

Sec61α constitutes the protein-conducting channel of the translocon complex crossing the membrane ten times, with both its N- and C-termini facing the cytosol. Viewed from the top, the protein adopts a square shape that can be divided in two pseudo-symmetric halves, the N-terminal halve containing

TMs 1-5 and the C-terminal comprising TMs 6-10 (red and blue TM segments in Fig 1, respectively). These two parts form an indentation in the centre through which the nascent chain passes and is aligned with the ribosomal exit tunnel [28]. From a lateral view, Sec61α has a rectangular contour and the channel within an hourglass shape [30]. In an inactive state the cytoplasmic entry to the channel has a diameter of approximately 20-25Å [24]. Close to the middle of the membrane the translocation pore reach its narrowest point (5-8Å) constituted by a ring of bulky hydrophobic residues followed by a short helix (TM 2a) that blocks the channel pore (Fig 1). After this "plug", the channel widens again towards the ER lumen. Nevertheless, it has been reported a significant increase in the pore diameter [31], which is probably needed to accommodate the multiple TM segments of multi-spanning nascent chains that may leave the translocon in pairs or groups (see below).

The β-subunit

The β-subunit is the smallest component of the Sec61 complex. It contains a single TM domain located next to TM segments 1 and 4 of Sec61α (Fig. 1A). Although this subunit is not essential either for translocation across ER membrane or insertion of TM segments into the lipid bilayer, it has been described to kinetically facilitate co-translational translocation [32], and to interact with the SR heterodimer probably facilitating recognition of unoccupied translocons by the RNC-SRP-SR complex [33]. The participation of Sec61β in the translocation process is also supported by its direct interaction with the nascent chain and the ribosome [34].

The γ-subunit

Sec61γ has two helices connected by an extended loop (Fig. 1). The first helical region, an amphipathic helix, sits parallel to the cytosolic side of the membrane and contacts with the cytoplasmic side of the Sec61α C-terminal halve. The second helix crosses the membrane diagonally interacting with both N- and C-terminal parts of Sec61α and acts as a clamp that brings both halves of Sec61α together [24].

Translocation and insertion of a nascent chain.

During co-translational insertion/translocation the nascent polypeptide is extruded into the translocon from the ribosome exit tunnel. The precise stoichiometry and structure of the actively engaged translocon-ribosome complex has been a great controversy over the past years. Initial cryo-EM studies indicated that 3-4 copies of the Sec61 complex could interact with the ribosome at the same time [35]. However, biochemical studies and the recent structures available strongly suggest that only one copy of the Sec61/SecY complex is required for translocation [24, 27-29, 36, 37]. Biochemical analysis of Sec61 point mutants [38], and the cryo-EM reconstructions of the ribosome-translocon pair indicate that the loops between TM segments 6-7 and 8-9 of the translocon are involved in this association [28, 39]. In fact, point mutations within those loops of the *E. coli* SecY are known to affect the ribosome-SecY interaction [39]. However, similar changes in loop L6 of the yeast translocon

did not affect binding to ribosome [28]. All in all indicates that, despite small differences, the ribosome-Sec junction is well conserved among species.

Although many details remain unknown, significant insight into the mechanism of membrane insertion has come from structural studies. The process starts with the engagement between the translocon complex and its cytosolic partner (that is, the ribosome in the co-translational pathway). Either this contact or the presence of the SS triggers the widening of the cytosolic side of the channel [25], including the hydrophobic ring that increases from ~5 to ~14Å [27]. In this pre-open state, displacement of TM segments 6, 8 and 9 from their position in the closed configuration would create a lateral "crack" between the two halves of Sec61α (*i.e.*, at the TM segments 2b and 7/8 interface), which would occur only in the cytosolic side of the channel. However, segment 2a retains its location keeping intact the permeability barrier. Once the SS enters into the channel as a loop, its first amino acids interact with the cytosolic residues of TM segment 8. At the same time, the hydrophobic core of the SS contacts TM segments 7 and 2b on both sides of the channel and with the phospholipids through the already open lateral crack [40]. As the elongation of the nascent chain continues two rearrangements occur in Sec61α. First, the plug should be displaced to leave room for the nascent polypeptide, which can now completely expand the channel. Second, the pairs formed by TM segments 2-3 on one side and 7-8 on the other halve move apart from each other (Fig. 1B) creating a lateral gate across the entire channel, which exposes the nascent polypeptide to the core of the membrane [27, 41]. The sequence within the translocon can then

partition into the lipids if it is hydrophobic enough (see TM domain requirements section), as the SS would do, or continue through the translocon into the ER lumen. The structural changes in the α-subunit are accompanied by a dramatic shift (Fig. 1B) in the location of the N-terminal helix form Sec61γ/SecE [27], which releases the clamp over Sec61α. Nevertheless, the opening of the lateral gate is not required to accommodate a translocating peptide within the channel [28]. Therefore, it is possible that the opening of the lateral gate is triggered by the presence of a TM segment inside the translocon, which would adjust its dynamic structure according to the nature of the polypeptide within the channel. During this process, the permeability barrier is kept by the coordinated *in* and *out* movement of the "plug" and the widening/narrowing of the hydrophobic ring, whilst the opening/closing of the lateral gate exposes hydrophobic segments to the lipid bilayer allowing their partition into the membrane.

TRAM

The TRanslocating chain-Associating Membrane protein (TRAM) was identified by crosslinking methods in reconstituted proteoliposomes [22]. Despite being recognized as an essential component for the translocation or insertion into the membrane of several secreted and membrane proteins, its precise function remains unknown. TRAM is an integral membrane protein with 8 TMs and both N- and C-termini facing the cytosol [42]. The role of TRAM in the translocation of secretory proteins is restricted to the insertion of the SS into the membrane [43], where TRAM has been found required for the

insertion of SS with either short hydrophobic sequences or with low overall hydrophobicity. Regarding the insertion of TM segments, TRAM has also been reported to crosslink with a wide variety of TM segments [44-48], some of them containing charged residues [49-51]. These observations together with the fact that TRAM itself contains an unusual high number of charged residues within its TM segments, led to the idea that TRAM could act as a chaperone for the integration of non-optimal TM segments by providing a more favorable context [42].

Translocon-associated proteins

Some other membrane proteins (*i.e.* TRAP, PAT-10, RAMP4 or BAP31) have been reported to interact with the translocon and modulate its function at some stage. However, their presence is not required for either insertion or translocation and thus they are not considered as a part of the translocon core complex.

The TRanslocon Associated Protein (TRAP) is a tetrameric complex (α, β, γ and δ) of integral membrane proteins [52]. It is associated with ribosome-Sec61 complexes with a 1:1 stoichiometry [29]. It has been proposed that TRAP facilitates the initiation of protein translocation [53] although the details of the mechanism remain unknown. PAT-10 was discovered as a translocon-associated protein when looking for Sec61 partners during opsin nascent chain insertion [50]. It is a membrane protein that cross-links with some of the opsin TM segments [54]. This interaction is independent of the presence of *N*-glycosylation sites, the amino acid sequence or the topology of its first TM

segment. Apparently, PAT-10 binding is triggered by the relative location of this TM segment within the opsin nascent chain. RAMP4 was also found tightly associated with the translocon [23]. RAMP4 is a small (66 amino acid) TA membrane protein implicated in promoting correct integration/folding of integral membrane proteins by facilitating subsequent glycosylation [55]. In a translating ribosome-translocon complex, RAMP4 is recruited to the Sec61 complex before the TM segment emerges from the ribosome exit tunnel; hence, it has been postulated that it is the presence of a TM sequence within the ribosome what triggers this recruitment [56]. Another protein that has been reported to interact with the translocon complex is BAP31. This multi-spanning integral membrane protein participates in the identification of misfolded proteins at the ER and their retro-translocation to the cytoplasm. The finding that BAP31 interacts both Sec61β and TRAM [57] suggests a role of the translocon in membrane protein quality control. The increasing amount of interacting partners of the translocon also indicates that different functions of the channel may be performed in association with different cellular components. Indeed, the Sec61 complex might just be the common player in a wide variety of transient complexes each one performing different but related functions.

Transmembrane domain requirements

Hydrophobicity

Individual TM helices follow an ordered insertion pathway, in which they pass from the tunnel in the large ribosomal subunit into the Sec61 translocon

channel and then exit the channel laterally into the surrounding lipids [30, 58]. Generally, the hydrophobicity of the TM sequence drives integration into the membrane. However, the insertion efficiency of TM segments by the translocon depends on amino acid composition, the positions of residues within the segment, TM segment orientation, and the helix length [59-62], suggesting that membrane insertion is fundamentally a fine-tuned thermodynamic partitioning process. Several TM segments from multi-spanning membrane proteins contain charged amino acids that are nevertheless tolerated in the membrane [63, 64]. Computational modeling suggests that integration of TM sequences with a central ionized residue might be assisted by helix–helix interactions within the membrane more than the stabilization of this ionized group by the translocon [65]. *In vivo* and *in vitro* studies suggest that the translocon may act as a facilitator in the insertion/selection process [59, 60, 66], where protein–lipid interactions "decide" the successful integration of the TM segment into the membrane through favorable acyl chain solvation [67] also affected by lipid composition [68]. Indeed, recent work in yeast showed that mutations in the hydrophobic constriction ring of Sec61p influence translocation efficiency modifying the hydrophobicity threshold for membrane insertion [69]. Such a mechanism based on lipid-mediated partitioning would accommodate the diversity of sequences that pass through the translocon in route to the membrane. Nevertheless, it has previously been suggested that the translocon complex can act as a chaperone during the integration of non-optimal TM segments. Indeed, a recent observation that ATP-depletion can halt TM segment release

from the translocon into the bilayer strongly supports this chaperone function [70], which supplement the thermodynamic partitioning process.

Amino acid preferences

A recent annotation on amino acid composition of α-helical TM segments showed that there is considerable information in sequence that relates to the intricate contacts between TM segments [71]. Indeed, there is a biased amino acid preferences depending on whether the residue is exposed to the lipid bilayer or to a soluble environment (Fig. 2). Using all annotation in the MPTopo database [72], we selected amino acids from TM segments and compared their occurrence to amino acids in non-TM segments. In total there were 206 proteins with known three-dimensional structure and topology, which had 1,244 TM segments. The total number of amino acids in TM segments was 25,281 compared to a total of 63,107 amino acids in non-TM regions. As previously reported [73] the hydrophobic residues Leu and Ala make up the bulk of the amino acids in the TM segments accounting for one fourth (24,5%) of all amino acids that are inserted through the translocon, but these two residues are also common in the non-TM regions (16,2%). This effect is even more evident for Gly since its prevalence is almost equal in TM or non-TM regions (Fig. 2). Interestingly, charged residues together with Pro are under-represented in TM domains relatively to non-TM regions. This feature is probably meaningful both in terms of hydrophobicity and helicity.

Helical conformation of transmembrane segments

Formation of a α-helix is critical for the membrane insertion of a TM segment. Even the most hydrophobic polypeptides could not insert into lipid bilayers without concomitant secondary structure formation [74]. One of the most intriguing challenges that membrane proteins have to face is desolvation and partitioning of the polar peptide bond from water into the membrane, which is as unfavorable as that of a charged side chain [75]. However, formation of intramolecular hydrogen bonds (*i.e.*, adoption of secondary structure) can compensate the loss of hydrogen bonds between the polypeptide backbone and water molecules (reviewed in [76]). Where does a predestined TM segment adopt its α-helical conformation? According to the two-stage model (see below), TM segments fold during insertion into the membrane and, in case of multi-spanning membrane proteins, before helix association [77]. However, some TM α-helices have been shown to fold already in the ribosomal tunnel [78-81] even before reaching the translocon or inserting into the lipid bilayer, which suggests that the folding inside the ribosome may regulate the fate of the nascent polypeptide.

Integration mechanism in multi-spanning membrane proteins

During biogenesis of multi-spanning membrane proteins, several TM segments in a single polypeptide need to be integrated by the Sec61□translocon. Unfortunately, our knowledge of the molecular mechanism underlying this process is still very limited. During translation and once the SS or a TM segment has reached the translocon, this first hydrophobic segment has to be relocated to accommodate the following TM

segment within the translocon pore. Whether at this point multiple TM segments partition into the membrane sequentially (*i.e.*, each TM segment exit the translocon individually [49]), or several TM segments can accumulate inside or in the proximities of the translocon and be released into the bilayer in pairs or groups [44, 50, 82, 83], is thought to be protein dependant. Recent structural data have shown that in the pre-open state the hydrophobic ring is widened to ~14Å in the direction of the lateral exit site [27], which is enough for accommodating more than one helix, especially because these dimensions could be further increased in a full open state [31]. It is also known that hydrophobic TM segments leave the translocon sequentially from the N-terminus to the C-terminus [82] and less hydrophobic segments interact with other TM segments at early stages of the membrane integration [46, 54, 84, 85]. More hydrophilic TM segments are forced by downstream hydrophobic sequence to adopt a TM disposition [86, 87]. However, whether these hydrophilic helices are spontaneously inserted or assisted by the Sec61 translocon to insert together with their partner helices is still unknown. Nevertheless, it has been suggested that inter-helical interactions are required to neutralize polar groups in TM sequences [76, 88]. Indeed, recent comparison of helix-helix interactions in available membrane protein structures reveal that they constitute one of the most distinctive characteristics of multi-spanning membrane proteins with more than 4 TM segments [89]. These helix-helix interactions might be coordinated *in vivo* by the translocon or its associated proteins. For example, TRAM (see above) plays a role assisting the integration of hydrophobic sequences containing charged

residues [43, 51]. Therefore, unraveling the functions of translocon-associated proteins will provide new insights into the integration mechanism of non-canonical TM segments.

Topology

During integration, nascent membrane proteins have to adopt the correct topology (*i.e.*, it has to define the number of TM segments and their orientation with respect to the plane of the lipid bilayer [90]), which is likely influenced by the translocon. However, whether a TM segment adopts a N-teminal cytosolic or reverse orientation depends on several factors. First, it has been observed that the folding state of an extra-membrane domain preceding a TM segment precludes its translocation and consequently forces the TM segment towards an N-terminal cytoplasmic orientation [91]. Second, the hydrophobicity of the TM sequence influences membrane orientation. For example, highly hydrophobic sequences promote N-terminal translocation despite the presence of moderate hydrophobic TM segments favor the opposite orientation [92]. Third, and most important, it has been long known that the distribution of charged residues between the flanking regions of a TM segment is a major determinant of topology in membrane proteins [93, 94]. The so-called 'positive-inside rule' was first observed for prokaryotic proteins, where bacteria maintain a net negative-inside electrical potential across the membrane, and a cytoplasmic bilayer leaflet enriched in negatively charged lipids also promotes charge bias. A similar skewed distribution was also identified later in eukaryotes [95], where the balance between positive and

negative charges drives protein topology. Indeed, changing the flanking charges by site-directed mutagenesis can reverse the topology of a TM segment (reviewed in [96]). Moreover, it has been recently demonstrated that certain residues of the translocon also contribute to the positive-inside orientation of signal sequences [97, 98]. Therefore, the amino acid sequence appears to be the primary determinant of final topology, which should be initially interpreted by the translocon. Nevertheless, it has also been reported that membrane lipid composition also influences the final topological orientation of membrane proteins (reviewed in [99]). In summary, both the amino acid sequence of a membrane protein and the collective determinants in the bilayer membrane influence protein topology.

Multi-spanning membrane proteins generally adopt their native orientation depending on the insertion of the SS or the first TM segment, which determines the alternant insertion of the rest of the protein. Nevertheless, drastic changes in loop regions that favor inverted orientations have only local effects [100]. Furthermore, its has been recently shown that the topology of a full length protein can be changed by simply adding a positively charged residue irrespectively of the region of the protein where the mutation was placed, including the C-terminal end of the protein [101]. Unfortunately, the molecular mechanisms by which downstream determinants contribute to the topology is yet unknown [102]. Therefore, experimental evidences are now challenging the classic static view for attaining membrane protein topology. For example, some proteins may adopt multiple topologies depending on the cellular localization or environment

[103] while others like viral membrane proteins have a strong preference for a specific topology [104].

Hydrophobic matching

The effect of the so-called hydrophobic matching on the assembly and orientation of TM segments has been widely studied [105]. A 'mismatch' happens when the hydrophobic thickness of the membrane do not match the length of the hydrophobic region of a TM segment [106]. There has been two types of hydrophobic mismatch described: (i) positive, when the membrane is not thick enough for a TM segment, and (ii) negative, when the length of the hydrophobic section of a TM segment is too short to span the hydrophobic core of the lipid bilayer. In both scenarios, either the membrane or the polypeptide will adapt to minimize the exposure of hydrophobic residues to the aqueous media (positive mismatch) or the extrusion of polar amino acids within the hydrophobic core of the membrane (negative mismatch) [107]. Both re-arrangements are known to be important for determining the final assembly of a membrane protein as proved by fluorescence [108-110] and chimeric overexpression of dimerizing TM segments in membrane-mimetic environments [111, 112]. The ability of the Sec61 translocon to handle negative mismatch has recently been studied [62]. In this work, it has been demonstrated that polyleucine segments as short as ~10 residues long integrate efficiently into the ER membrane. Finally, hydrophobic matching may reflect an evolutionary strategy to regulate the activity of membrane

proteins by allowing the adaptation of TM segment lengths to bilayer thickness in different cellular membranes [113].

Folding and assembly of multi-spanning membrane proteins

Forces behind the folding of membrane proteins

Next, we briefly introduce the molecular interactions driving protein folding within membranes. For a recent complete review see [74, 75]. While hydrophobic collapse is a major driving force in the folding of soluble proteins, its role in membrane proteins is mostly limited to the formation of secondary structures across the lipid bilayer. Similarly, salt bridges and aromatic interactions do not have a weighty contribution in membrane protein folding. Conversely, inter-helical hydrogen bonding [114, 115] and specially van der Waals forces have been identified as major promoters of membrane protein folding [116, 117]. Therefore, the restrictions imposed by the lipid bilayer allow for an effective folding of TM segments of integral membrane proteins despite the low contribution of hydrophobic forces and the reduced effect of salt bridges and aromatic interactions [118].

Folding and assembly of membrane proteins: The two-stage model.

The folding and assembly of helical membrane proteins was schematized more than two decades ago as a two-stage process [77]. First, each TM helix is formed and independently inserted into the lipid bilayer. Second, these helices interact with each other to establish the final structure of the protein.

Although this simplified view has been refined in the past years still constitutes a valid conceptual approach.

In vivo, the insertion into the ER membrane occurs co-translationally via the translocon complex. In this scenario a TM segment does not insert into the membrane spontaneously, instead the translocon facilitates its partition from the aqueous environment within the translocon pore into the lipid bilayer. After insertion, or for some proteins during insertion, the TM helices interact with each other to form higher order structures. These interactions create a microenvironment that permits further changes in the protein structure, such as insertion into the membrane of re-entrant loops or short polypeptides, membrane packing of non-α-helical segments and binding of prosthetics groups [119].

Finally, the influence of the specific lipid environment during the assembly of TM segments should also be taken into account. The lipid and protein components of biological membranes have co-evolved allowing membrane proteins to assemble and function in the heterogenic environment provided by the diverse lipid bilayers in a cell. Not only membrane thickness (see Hydrophobic matching section) but also membrane lateral pressure [120], charge density [121], and even unique lipid-protein interactions [122] have been identified as structural determinants of membrane proteins. Furthermore, very recent cryo-EM studies using RNC complexes bound to SecY reconstituted in nano-discs revealed an interaction of the ribosome with lipids, leading to a disorder in the lipid microenvironment adjacent to the translocon, which may contribute to favor membrane insertion of TM

segments [123]. All in all, the final structure of a multi-spanning membrane protein will not solely be defined by protein-protein and lipid-protein interactions but also by the folding of its soluble domains. Thus, the aqueous environment on both sides of the membrane imposes restrictions to the folding of the extra-membranous regions, and by extension to the overall protein structure.

Concluding Remarks

Membrane protein integration appears to be orchestrated by multiple determinants and factors that in unlimited combinations give raise to native protein structures. During protein targeting, TM segment insertion and assembly into the membrane several inter-connected processes occur simultaneously. Structural studies of the translocon together with *in vitro* quantitative thermodynamic analyses and biophysical dissection of TM interactions have resulted in significant advance of our understanding of membrane protein integration into the lipid bilayer. Our current knowledge coupled with bioinformatics analysis [124] is opening now opportunities for *de novo* membrane protein structure prediction and design.

Acknowledgments

The authors acknowledge financial support from the Spanish Ministry of Science and Innovation (BFU2009-08401/BMC to I.M. and BFU2010-

19310/BMC to M.A.M-R), and from the Generalitat Valenciana (PROMETEO/2010/005 and ACOMP/2011/025 to I.M., and ACOMP/2011/048 to M.A.M-R.).

Figure legends

Fig. 1. Translocon Structure. Top view of the translocon structure. A) Closed structure of the translocon from *M. jannaschii* (PDB ID code 1RHZ) [20]. B) Partially open structure of the translocon from *P. furiosus* (PDB ID code 3MP7) [23]. In both panels all TM segments of Sec61α are colored (red and blue for each half, see text) except for the β and γ subunits, which are shown in grey. All TM segments are numbered for easy comparison between the open and closed structures. Dotted arrows in panel B indicate helix displacements required for the widening of the channel and opening of the lateral gate. A solid arrow shows the lateral gate exit pathway of a TM segment from the interior of the channel into the membrane.

Fig. 2. Amino acid preferences in TM segments compared to loop regions (non-TM) in membrane protein structures. Top two rows show the percentage of occurrence of all amino acid types in TM segments and non-TM segments in membrane proteins of known structure. Lower plot shows the log odds ratio of the occurrence. Briefly, a log odd ratio the is the $\log_{10}$ ratio of the odds of an amino acid occurring in TM segments to the odds of it occurring in a non-TM segment. Positive log odds indicate over-occurrence of the amino acid type in TM segments. Negative log odds indicate under-representation of the amino acid type in TM segments. Amino acids are colored according to an arbitrary division of their log odds (*i.e.*, green color for log odds > 0.3; orange color for 0.3 • log odds • -0.3; and red color for log odds < -0.3).

References

1. Wallin E & von Heijne G (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci* **7**, 1029-1038.
2. Egea PF, Stroud RM & Walter P (2005) Targeting proteins to membranes: structure of the signal recognition particle. *Current Opinion in Structural Biology* **15**, 213-220.
3. Johnson AE & van Waes MA (1999) The translocon: A dynamic gateway at the ER membrane. *Annual Review of Cell and Developmental Biology* **15**, 799-842.
4. Rapoport TA (2008) Protein transport across the endoplasmic reticulum membrane. *FEBS J* **275**, 4471-4478.
5. Ng DTW, Brown JD & Walter P (1996) Signal sequences specify the targeting route to the endoplasmic reticulum. *J Cell Biol* **134**, 269-278.
6. Cross BC, Sinning I, Luirink J & High S (2009) Delivering proteins for export from the cytosol. *Nat Rev Mol Cell Biol* **10**, 255-264.
7. Pool MR (2005) Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). *Mol Membr Biol* **22**, 3-15.
8. Halic M, Blau M, Becker T, Mielke T, Pool MR, Wild K, Sinning I & Beckmann R (2006) Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444**, 507-511.
9. Lakkaraju AKK, Mary C, Scherrer A, Johnson AE & Strub K (2008) SRP Keeps Polypeptides Translocation-Competent by Slowing Translation to Match Limiting ER-Targeting Sites. *Cell* **133**, 440-451.
10. Gilmore R, Blobel G & Walter P (1982) Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *J Cell Biol* **95**, 463-469.
11. Montoya G, Svensson C, Luirink J & Sinning I (1997) Crystal structure of the NG domain from the signal recognition particle receptor FtsY. *Nature* **385**, 365-368.
12. Song W, Raden D, Mandon EC & Gilmore R (2000) Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell* **100**, 333-343.
13. Halic M, Gartmann M, Schlenker O, Mielke T, Pool MR, Sinning I & Beckmann R (2006) Signal Recognition Particle Receptor Exposes the Ribosomal Translocon Binding Site. *Science* **312**, 745-747, doi: 10.1126/science.1124864.
14. Janda CY, Li J, Oubridge C, Hernandez H, Robinson CV & Nagai K (2010) Recognition of a signal peptide by the signal recognition particle. *Nature* **465**, 507-510.
15. Ataide SF, Schmitz N, Shen K, Ke A, Shan S-o, Doudna JA & Ban N (2011) The Crystal Structure of the Signal Recognition Particle in Complex with Its Receptor. *Science* **331**, 881-886.
16. Hainzl T, Huang S, Merilainen G, Brannstrom K & Sauer-Eriksson AE (2011) Structural basis of signal-sequence recognition by the signal recognition particle. *Nat Struct Mol Biol* **18**, 389-391.
17. 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.
18. Becker J, Walter W, Yan W & Craig EA (1996) Functional interaction of cytosolic hsp70 and a DnaJ- related protein, Ydj1p, in protein translocation in vivo. *Mol Cell Biol* **16**, 4378-4386.

19. Wilkinson BM, Regnacq M & Stirling CJ (1997) Protein translocation across the membrane of the endoplasmic reticulum. *J Membr Biol* **155**, 189-197.
20. Borgese N & Fasana E (2011) Targeting pathways of C-tail-anchored proteins. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 937-946.
21. Rabu C, Schmid V, Schwappach B & High S (2009) Biogenesis of tail-anchored proteins: the beginning for the end? *J Cell Sci* **122**, 3605-3612.
22. Görlich D, Hartmann E, Prehn S & Rapoport TA (1992) A Protein of the Endoplasmic Reticulum Involved Early in Polypeptide Translocation. *Nature* **357**, 47-52.
23. Görlich D & Rapoport TA (1993) Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75**, 615-630.
24. Van den Berg B, Clemons WM, Jr., Collinson I, Modis Y, Hartmann E, Harrison SC & Rapoport TA (2004) X-ray structure of a protein-conducting channel. *Nature* **427**, 36-44.
25. Tsukazaki T, Mori H, Fukai S, Ishitani R, Mori T, Dohmae N, Perederina A, Sugita Y, Vassylyev DG, Ito K, et al. (2008) Conformational transition of Sec machinery inferred from bacterial SecYE structures. *Nature* **455**, 988-991.
26. Zimmer J, Nam Y & Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**, 936-943.
27. Egea PF & Stroud RM (2010) Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc Natl Acad Sci U S A* **107**, 17182-17187.
28. Becker T, Bhushan S, Jarasch A, Armache JP, Funes S, Jossinet F, Gumbart J, Mielke T, Berninghausen O, Schulten K, et al. (2009) Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369-1373.
29. Menetret JF, Hegde RS, Aguiar M, Gygi SP, Park E, Rapoport TA & Akey CW (2008) Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure* **16**, 1126-1137.
30. Rapoport TA, Goder V, Heinrich SU & Matlack KE (2004) Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol* **14**, 568-575.
31. Hamman BD, Chen J-C, Johnson EE & Johnson AE (1997) The aqueous pore through the translocon has a diameter of 40-60 Å during cotrsnaltional protein translocation at the ER membrane. *Cell* **89**, 535-544.
32. Kalies KU, Rapoport TA & Hartmann E (1998) The beta subunit of the Sec61 complex facilitates cotranslational protein transport and interacts with the signal peptidase during translocation. *J Cell Biol* **141**, 887-894.
33. Jiang Y, Cheng Z, Mandon EC & Gilmore R (2008) An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol* **180**, 1149-1161, doi: jcb.200707196 [pii] 10.1083/jcb.200707196.
34. Levy R, Wiedmann M & Kreibich G (2001) In vitro binding of ribosomes to the beta subunit of the Sec61p protein translocation complex. *J Biol Chem* **276**, 2340-2346.
35. Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G & Frank J (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex [see comments]. *Science* **278**, 2123-2126.
36. Yahr TL & Wickner WT (2000) Evaluating the oligomeric state of SecYEG in preprotein translocase. *EMBO J* **19**, 4393-4401.

37. Cannon KS, Or E, Clemons WM, Jr., Shibata Y & Rapoport TA (2005) Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J Cell Biol* **169**, 219-225.
38. Cheng Z, Jiang Y, Mandon EC & Gilmore R (2005) Identification of cytoplasmic residues of Sec61p involved in ribosome binding and cotranslational translocation. *J Cell Biol* **168**, 67-77.
39. Menetret JF, Schaletzky J, Clemons WM, Jr., Osborne AR, Skanland SS, Denison C, Gygi SP, Kirkpatrick DS, Park E, Ludtke SJ, et al. (2007) Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol Cell* **28**, 1083-1092.
40. Plath K, Mothes W, Wilkinson BM, Stirling CJ & Rapoport TA (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94**, 795-807.
41. Martoglio B, Hofmann MW, Brunner J & Dobberstein B (1995) The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell* **81**, 207-214.
42. Tamborero S, Vilar M, Martinez-Gil L, Johnson AE & Mingarro I (2011) Membrane Insertion and Topology of the Translocating Chain-Associating Membrane Protein (TRAM). *J Mol Biol* **406**, 571-582.
43. Voigt S, Jungnickel B, Hartmann E & Rapoport TA (1996) Signal sequence-dependent function of the TRAM protein during early pphases of protein transport across the endoplasmic reticulum membrane. *J Cell Biol* **134**, 25-35.
44. McCormick PJ, Miao Y, Shao Y, Lin J & Johnson AE (2003) Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. *Mol Cell* **12**, 329-341.
45. Do H, Falcone D, Lin J, Andrews DW & Johnson AE (1996) The cotransational integration of membrane proteins into the phospholipid bilayer is a multistep process. *Cell* **85**, 369-378.
46. Heinrich SU & Rapoport TA (2003) Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane. *Embo J* **22**, 3654-3663.
47. Sauri A, McCormick PJ, Johnson AE & Mingarro I (2007) Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. *J Mol Biol* **366**, 366-374.
48. Martinez-Gil L, Johnson AE & Mingarro I (2010) Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *J Virol* **84**, 5520-5527.
49. Heinrich SU, Mothes W, Brunner J & Rapoport TA (2000) The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**, 233-244.
50. Meacock SL, Lecomte FJ, Crawshaw SG & High S (2002) Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. *Mol Biol Cell* **13**, 4114-4129.
51. Cross BC & High S (2009) Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J Cell Sci* **122**, 1768-1777.
52. Hartmann E, Gorlich D, Kostka S, Otto A, Kraft R, Knespel S, Burger E, Rapoport TA & Prehn S (1993) A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur J Biochem* **214**, 375-381.
53. Fons RD, Bogert BA & Hegde RS (2003) Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J Cell Biol* **160**, 529-539.

54. Ismail N, Crawshaw SG, Cross BC, Haagsma AC & High S (2008) Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. *Biochem J* **411**, 495-506.
55. Yamaguchi A, Hori O, Stern DM, Hartmann E, Ogawa S & Tohyama M (1999) Stress-Associated Endoplasmic Reticulum Protein 1 (Serp1)/Ribosome-Associated Membrane Protein 4 (Ramp4) Stabilizes Membrane Proteins during Stress and Facilitates Subsequent Glycosylation. *The Journal of Cell Biology* **147**, 1195-1204.
56. Pool MR (2009) A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *J Cell Biol* **185**, 889-902.
57. Wang B, Heath-Engel H, Zhang D, Nguyen N, Thomas DY, Hanrahan JW & Shore GC (2008) BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTRDeltaF508 via the derlin-1 complex. *Cell* **133**, 1080-1092.
58. Alder NN & Johnson AE (2004) Cotranslational membrane protein biogenesis at the endoplasmic reticulum. *J Biol Chem* **279**, 22787-22790.
59. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH & von Heijne G (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-381.
60. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH & von Heijne G (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026-1030.
61. Lundin C, Kim H, Nilsson I, White SH & von Heijne G (2008) Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for Nin/Cout and Nout/Cin transmembrane helices. *Proceedings of the National Academy of Sciences* **105**, 15702-15707.
62. Jaud S, Fernandez-Vidal M, Nilsson I, Meindl-Beinker NM, Hubner NC, Tobias DJ, von Heijne G & White SH (2009) Insertion of short transmembrane helices by the Sec61 translocon. *Proc Natl Acad Sci U S A* **106**, 11588-11593.
63. Martinez-Gil L, Perez-Gil J & Mingarro I (2008) The surfactant peptide KL4 sequence is inserted with a transmembrane orientation into the endoplasmic reticulum membrane. *Biophys J* **95**, L36-38.
64. White SH & von Heijne G (2008) How translocons select transmembrane helices. *Annu Rev Biophys* **37**, 23-42.
65. Rychkova A, Vicatos S & Warshel A (2010) On the energetics of translocon-assisted insertion of charged transmembrane helices into membranes. *Proceedings of the National Academy of Sciences* **107**, 17598-17603.
66. Mulvihill CM & Deber CM (2010) Evidence that the translocon may function as a hydropathy partitioning filter. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1798**, 1995-1998.
67. Johansson AC & Lindahl E (2009) Titratable amino acid solvation in lipid membranes as a function of protonation state. *J Phys Chem B* **113**, 245-253.
68. Johansson AC & Lindahl E (2009) The role of lipid composition for insertion and stabilization of amino acids in membranes. *J Chem Phys* **130**, 185101.
69. 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.
70. Pitonzo D, Yang Z, Matsumura Y, Johnson AE & Skach WR (2009) Sequence-Specific Retention and Regulated Integration of a Nascent Membrane Protein by the ER Sec61 Translocon. *Mol Biol Cell* **20**, 685-698.

71. Arce J, Sturgis JN & Duneau JP (2009) Dissecting membrane protein architecture: An annotation of structural complexity. *Biopolymers* **91**, 815-829.
72. Jayasinghe S, Hristova K & White SH (2001) MPtopo: A database of membrane protein topology. *Protein Sci* **10**, 455-458.
73. Ulmschneider MB & Sansom MS (2001) Amino acid distributions in integral membrane protein structures. *Biochim Biophys Acta* **1512**, 1-14.
74. Fiedler S, Broecker J & Keller S (2010) Protein folding in membranes. *Cellular and Molecular Life Sciences* **67**, 1779-1798.
75. White SH & Wimley WC (1999) Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* **28**, 319-365.
76. Bowie JU (2011) Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol* **21**, 42-49.
77. Popot JL & Engelman DM (1990) Membrane protein folding and oligomerization - The 2-stage model. *Biochemistry* **29**, 4031-4037.
78. Mingarro I, Nilsson I, Whitley P & von Heijne G (2000) Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC Cell Biol* **1**, 3.
79. Woolhead CA, McCormick PJ & Johnson AE (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725-736.
80. Tu LW & Deutsch C A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation. *Journal of Molecular Biology* **396**, 1346-1360.
81. Bhushan S, Gartmann M, Halic M, Armache JP, Jarasch A, Mielke T, Berninghausen O, Wilson DN & Beckmann R (2010) alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat Struct Mol Biol* **17**, 313-317.
82. Sadlish H, Pitonzo D, Johnson AE & Skach WR (2005) Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. *Nat Struct Mol Biol* **12**, 870-878.
83. Sauri A, Saksena S, Salgado J, Johnson AE & Mingarro I (2005) Double-spanning Plant Viral Movement Protein Integration into the Endoplasmic Reticulum Membrane Is Signal Recognition Particle-dependent, Translocon-mediated, and Concerted. *J Biol Chem* **280**, 25907-25912.
84. Ota K, Sakaguchi M, Hamasaki N & Mihara K (2000) Membrane integration of the second transmembrane segment of band 3 requires a closely apposed preceding signal-anchor sequence. *J Biol Chem* **275**, 29743-29748.
85. Hedin LE, Ojemalm K, Bernsel A, Hennerdal A, Illergard K, Enquist K, Kauko A, Cristobal S, von Heijne G, Lerch-Bader M, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J Mol Biol* **396**, 221-229.
86. Ota K, Sakaguchi M, von Heijne G, Hamasaki N & Mihara K (1998) Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. *Mol Cell* **2**, 495-503.
87. Kida Y, Morimoto F & Sakaguchi M (2007) Two translocating hydrophilic segments of a nascent chain span the ER membrane during multispanning protein topogenesis. *J Cell Biol* **179**, 1441-1452.
88. Hermansson M & von Heijne G (2003) Inter-helical Hydrogen Bond Formation During Membrane Protein Integration into the ER Membrane. *Journal of Molecular Biology* **334**, 803-809.

89. Fuchs A & Frishman D (2010) Structural comparison and classification of alpha-helical transmembrane domains based on helix interaction patterns. *Proteins: Structure, Function, and Bioinformatics* **78**, 2587-2599, doi: 10.1002/prot.22768.
90. von Heijne G (2006) Membrane-protein topology. *Nat Rev Mol Cell Biol* **7**, 909-918.
91. 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.
92. Goder V & Spiess M (2003) Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO J* **22**, 3645-3653.
93. von Heijne G (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456-458.
94. 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.
95. Hartmann E, Rapoport TA & Lodish HF (1989) Predicting the orientation of eukaryotic membrane proteins. *Proc Natl Acad Sci USA* **86**, 5786-5790.
96. Goder V & Spiess M (2001) Topogenesis of membrane proteins: determinants and dynamics. *FEBS Lett* **504**, 87-93.
97. 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.
98. 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.
99. Dowhan W & Bogdanov M (2009) Lipid-dependent membrane protein topogenesis. *Annu Rev Biochem* **78**, 515-540.
100. 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.
101. Seppala S, Slusky JS, Lloris-Garcera P, Rapp M & von Heijne G (2010) Control of membrane protein topology by a single C-terminal residue. *Science* **328**, 1698-1700.
102. Nilsson I, Witt S, Kiefer H, Mingarro I & von Heijne G (2000) Distant Downstream Sequence Determinants Can Control N-tail Translocation during Protein Insertion into the Endoplasmic Reticulum Membrane. *J Biol Chem* **275**, 6207-6213.
103. Hegde RS, Voigt S & Lingappa VR (1998) Regulation of protein topology by trans-acting factors at the endoplasmic reticulum. *Mol Cell* **2**, 85-91.
104. Sauri A, Tamborero S, Martinez-Gil L, Johnson AE & Mingarro I (2009) Viral membrane protein topology is dictated by multiple determinants in its sequence. *J Mol Biol* **387**, 113-128.
105. Dumas F, Lebrun MC & Tocanne JF (1999) Is the protein/lipid hydrophobic matching principle relevant to membrane organization and functions? *FEBS Lett* **458**, 271-277.
106. Killian JA (1998) Hydrophobic mismatch between proteins and lipids in membranes. *Biochim Biophys Acta* **10**, 401-415.
107. Andersen OS & Koeppe RE, 2nd (2007) Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct* **36**, 107-130, doi: 10.1146/annurev.biophys.36.040306.132643.

108. Sparr E, Ash WL, Nazarov PV, Rijkers DTS, Hemminga MA, Tieleman DP & Killian JA (2005) Self-association of Transmembrane Œ±-Helices in Model Membranes. *Journal of Biological Chemistry* **280**, 39324-39331.
109. Mall S, Broadbridge R, Sharma RP, East JM & Lee AG (2001) Self-association of model transmembrane alpha-helices is modulated by lipid structure. *Biochemistry* **40**, 12379-12386.
110. Ren J, Lew S, Wang J & London E (1999) Control of the Transmembrane Orientation and Interhelical Interactions within Membranes by Hydrophobic Helix Length. *Biochemistry* **38**, 5905-5912.
111. Orzaez M, Perez-Paya E & Mingarro I (2000) Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. *Protein Sci* **9**, 1246-1253.
112. Orzaez M, Lukovic D, Abad C, Perez-Paya E & Mingarro I (2005) Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. *FEBS Lett* **579**, 1633-1638.
113. Sharpe HJ, Stevens TJ & Munro S (2010) A comprehensive comparison of transmembrane domains reveals organelle-specific properties. *Cell* **142**, 158-169.
114. Joh NH, Min A, Faham S, Whitelegge JP, Yang D, Woods VL & Bowie JU (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266-1270.
115. Popot JL & Engelman DM (2000) Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* **69**, 881-922.
116. MacKenzie KR, Prestegard JH & Engelman DM (1997) A transmembrane helix dimer: Structure and implications. *Science* **276**, 131-133.
117. Eilers M, Shekar SC, Shieh T, Smith SO & Fleming PJ (2000) Internal packing of helical membrane proteins. *Proc Natl Acad Sci U S A* **97**, 5796-5801.
118. Bowie JU (2005) Solving the membrane protein folding problem. *Nature* **438**, 581-589.
119. Engelman DM, Chen Y, Chin CN, Curran AR, Dixon AM, Dupuy AD, Lee AS, Lehnert U, Matthews EE, Reshetnyak YK, et al. (2003) Membrane protein folding: beyond the two stage model. *FEBS Lett* **555**, 122-125.
120. van den Brink-van der Laan E, Chupin V, Killian JA & de Kruijff B (2004) Small Alcohols Destabilize the KcsA Tetramer via Their Effect on the Membrane Lateral Pressure. *Biochemistry* **43**, 5937-5942.
121. Bogdanov M & Dowhan W (1999) Lipid-assisted Protein Folding. *Journal of Biological Chemistry* **274**, 36827-36830, doi: 10.1074/jbc.274.52.36827.
122. Lee AG (2003) Lipid-protein interactions in biological membranes: a structural perspective. *Biochim Biophys Acta* **1612**, 1-40.
123. Frauenfeld J, Gumbart J, Sluis EO, Funes S, Gartmann M, Beatrix B, Mielke T, Berninghausen O, Becker T, Schulten K, et al. (2011) Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. *Nat Struct Mol Biol* **18**, 614-621.
124. Elofsson A & von Heijne G (2007) Membrane protein structure: prediction versus reality. *Annu Rev Biochem* **76**, 125-140.