Universitat de València
Departament de Bioquímica i Biologia Molecular

Manuel Bañó Polo

TESIS DOCTORAL
2013

TRANSMEMBRANE HELICES: INSERTION AND ASSEMBLY INTO BIOLOGICAL MEMBRANES

Tesis dirigida por el Dr. Ismael Mingarro

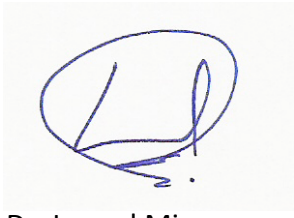
Memoria presentada para optar al título de Doctor Internacional por la Universitat de València

Programa de Doctorado 30G Señalización celular y patologias asociadas

Ismael MINGARRO MUÑOZ, Catedràtic del Departament de Bioquímica i Biologia Molecular de la Universitat de València,

MANIFESTA que En Manuel BAÑÓ POLO, Llicenciat en Bioquímica per la Universitat de València i en Química per la Universitat Jaume I, ha realitzat sota la meua direcció al Laboratori de Proteïnes de Membrana de la ERI BioTecMed (Departament de Bioquímica i Biologia Molecular) la memòria titulada "*Transmembrane helices: insertion and assembly into biological membranes*" que presenta per a optar al grau de Doctor.

I perquè així conste i tinga els efectes oportuns, signe el present document en Burjassot el 17 de maig de 2013



Dr. Ismael Mingarro
Director de la Tesi

Para la realización de esta tesis doctoral, Manuel Bañó Polo ha disfrutado de una Beca del Programa de Formación de Profesorado Universitario (FPU, Convocatoria 2008) y de una Ayuda para la Realización de Estancia Breve en la Texas A&M University dentro del Programa de Formación de Profesorado Universitario (Ayuda Estancias Breves 2011), habiendo sido ambas ayudas concedidas por el Ministerio de Educación, Cultura y Deporte.

El trabajo se ha enmarcado dentro de los proyectos: "Inserción, ensamblaje y topologia de proteínas de membrana" (BFU2006-08542) financiado por el Ministerio de Educación y Ciencia; "Membrane protein expression, folding and dynamics" (BFU2009-08401/BMC) financiado por el Ministerio de Ciencia e Innovación, y "Membrane Protein Biology" (BFU2012-39482) financiado por el Ministerio de Economía y Competitividad, de los que el director de esta Tesis es Investigador Principal.

La Tesis ha sido realizada en el "Membrane Proteins Laboratory" (MemProtLab) del Departament de Bioquímica i Biologia Molecular de la Universitat de València y parcialmente en "Department of Molecular and Cellular Medicine, Health Science Center" de la Universidad de Texas A&M

Abbreviations:

ATP:	Adenosine triphosphate
bR:	Bacteriorhodopsin
Cryo-EM:	Cryo-electron microscopy
Ct	C-terminus
DOPC:	1,2-dioleoyl-sn-glycero-3-phosphocholine
ER:	Endoplasmatic reticulum
GpA:	Glycophorin A
HR:	Hydrophobic region
IMP:	Integral membrane protein
Lep:	Leader peptidase
MP:	Membrane Protein
OST	Oligosaccharyl transferase
PE:	Phosphatidylethanolamine
PMP:	Peroxisomal membrane protein
PV:	Poliovirus
RM:	Rough microsomes
RNC:	Ribosome nascent chain complex
SDS-PAGE:	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNARE:	Soluble NSF Attachment Protein Receptor
SR:	Signal recognition particle receptor
SRP:	Signal recognition particle
SS:	Signal sequence
TM:	Transmembrane
TMD:	Transmembrane domain
TMH:	Transmembrane helix
TRAM:	Translocating chain-associating membrane protein

Amino acids, one and three letter codes:

Any amino acid	Xaa	X		Leucine	Leu	L	
Alanine	Ala	A		Lysine	Lys	K	
Arginine	Arg	R		Methionine	Met	M	
Asparagine	Asn	N		Phenylalanine	Phe	F	
Aspartic Acid	Asp	D		Proline	Pro	P	
Cysteine	Cys	C		Serine	Ser	S	
Glutamine	Gln	Q		Threonine	Thr	T	
Glutamic acid	Glu	E		Tryptophan	Trp	W	
Glycine	Gly	G		Tyrosine	Tyr	Y	
Histidine	His	H		Valine	Val	V	
Isoleucine	Iso	I					

INTRODUCTION

Most of the processes in a living cell are carried out by proteins. Depending on the needs of the cell, different proteins will interact and form the molecular machines demanded for the moment. A subset of proteins called integral membrane proteins are responsible for the interchange of matter and information across the biological membrane, the lipid bilayer surrounding and defining the cell. Molecular traffic and information flow across biological membranes are mediated by the channel, transporter, and receptor activities of the proteins embedded within them. These proteins are characterized by the presence of transmembrane domains (TMDs), polypeptide sequences uniquely adapted to insert, fold, and function in the complex solvent environment of cellular membranes. TMDs may be divided into two broad classes based on their secondary structure (**Figure 1**): β-barrels that are resident in the outer membranes of bacteria, mitochondria, and chloroplasts and α-helical bundles that traverse the cytoplasmic membranes of archaeal, bacterial, and eukaryotic cells.

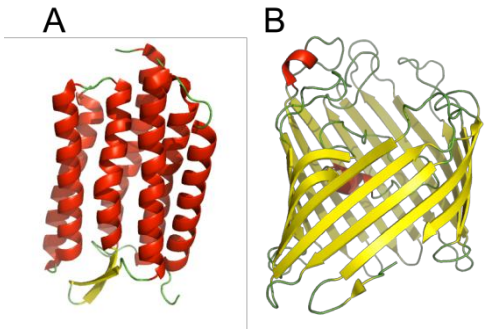


Figure 1. The two major structural types of membrane proteins, cartoon representation. (A) Helical bundle. Bacteriorhodopsin, PDB code 1NTU. (B) β-barrel. Phosphoporin, PDB code 1PHO. Helical regions and β-sheets are shown in red and yellow respectively.

In this thesis we focus on the α-helical membrane proteins, a class of molecules that represents 20–30% of all open reading frames in fully sequenced genomes (Krogh et al., 2001). However, our understanding of their biosynthesis and folding lags far behind our understanding of water-soluble proteins, due to the complexity of their purification and membrane characteristics.

I.1. Biological membranes

Cells are the basic unit of life, defined and delimited by their biological membranes. All eukaryotic cells have organelles, internal compartments that are enclosed by membranes, which specialize in different cellular functions.

i.1.1. Basic structure of phospholipid membranes

Phospholipids in biological membranes are amphipathic molecules. They consist of a hydrophobic hydrocarbon tail and a hydrophilic phosphate-containing headgroup (**Figure 2**A). The hydrophobic effect will cause them to form bilayers, consisting of two sheets of lipids with their hydrophobic tails facing each other, see **Figure 2**B. The hydrophilic headgroups shield the tails, forming an interface between the hydrophobic core of the membrane and the aqueous surroundings (Granseth et al., 2005). There is no sharp border between the hydrophobic core and the surrounding water as this interface region provides a zone of gradually changing hydrophobicity, see **Figure 2**C. In 1992, Wiener and White determined the structure of a bilayer of pure DOPC lipids (Wiener and White, 1992). The hydrophobic core was about 30 Å thick, while the interface region extended about 15 Å on either side, see **Figure 2**C. The tails of the lipids can be of different lengths and stiffnesses and the size and electrical charge of the headgroup can differ. Steroidic lipids like cholesterol are very rigid and confer stability (Nelson et al., 2008). The features and properties of a biological membrane, thickness, fluidity, curvature, pressure, etc., will vary with the lipid composition (van Meer et al., 2008). It varies from organism to organism and between organelles in the same cell, even between leaflets of the same bilayer.

Considerations based on the ~30-35 Å hydrophobic thickness of a typical membrane bilayer and on α-helix geometry suggested that a TMD sequence would comprise a continuous sequence of ~20-23 mainly hydrophobic residues along with some polar, and a few ionic, side chains where required for function. Analyses of the increasing structural database of helical TMDs remain remarkably consistent with initial predictions. TMDs average ~24.0 ± 5.6 residues in length and exhibit a hydrophobicity gradient in their side chain distribution, with a highly apolar central region flanked by more polar and charged residues (Baeza-Delgado et al., 2013).

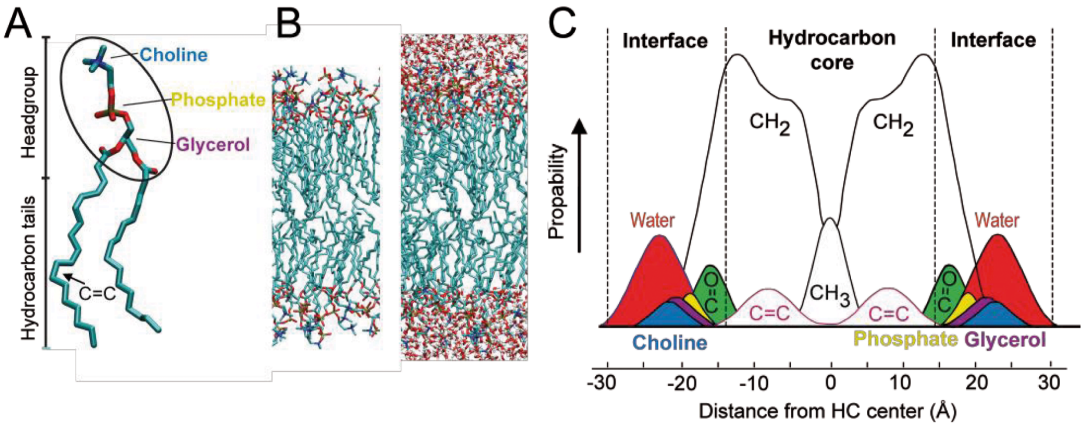


Figure 2. Phospholipids and membranes (A) A ribbon representation of a phosphatidylcholine lipid with a C17- and a C14- hydrocarbon tail. Cyan, carbon; red, oxygen; blue, nitrogen; yellow, phosphate. Hydrogens are omitted for clarity. (B) A simulated phospholipid bilayer. Molecular groups colored as in A. Without (left) and with (right) surrounding water molecules. (C) The structure of a DOPC- bilayer along the membrane normal, shown as the probability of finding different chemical groups at a given distance from the center of the hydrocarbon core. Adapted from (White and Wimley, 1999)

i.1.2. Biological membranes consist of more than lipids

One of the modern models of a cell membrane depicted it as an ocean of lipids with proteins floating around in it, the fluid mosaic model (Singer and Nicolson, 1972). The truth has been shown to be somewhat different (Jacobson et al., 2007; Vereb et al., 2003). As shown in **Figure 3**, in some membranes, the lipids act more as a cohesive moiety between proteins with up to 30 000 molecules of integrated membrane proteins per μm^2 membrane (Dupuy and Engelman, 2008; Quinn et al., 1984; Roof and Heuser, 1982). So, the current view shows a more mosaic than fluid model (**Figure 3**).

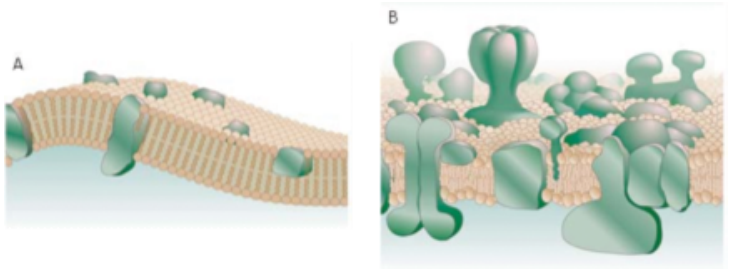


Figure 3. General models for membrane structure. (A) The Singer–Nicholson 'fluid mosaic model'. (B) An amended and updated version. Taken from (Engelman, 2005)

I.2. Biogenesis and bilayer integration of α-helical membrane proteins

i.2.1. Insertion of α-helical membrane proteins

In the cell, the vast majority of proteins are synthesized in the cytosol, where the ribosomes translate mRNA codons into amino acids and add them to the growing polypeptide chain. Soluble cytosolic proteins simply fold as they emerge from the ribosome, in some cases assisted by molecular chaperones. For proteins meant for secretion or integration into a lipid bilayer, the process is more complicated due to the presence of the membrane barrier and have been described both co-translational and post-translational pathways for translocation across (secreted proteins) or insertion into the membrane (integral membrane proteins).

The most primary and well-studied system is the SRP-dependent secretase (Sec) pathway. Its major components are the signal recognition particle (SRP), the signal recognition particle receptor (SR) and the core of the Sec translocon, which forms a pore through the membrane, reviewed in (Rapoport, 2007).

After translation is initiated, there are two major steps in membrane and secreted protein biogenesis, see **Figure 4**:

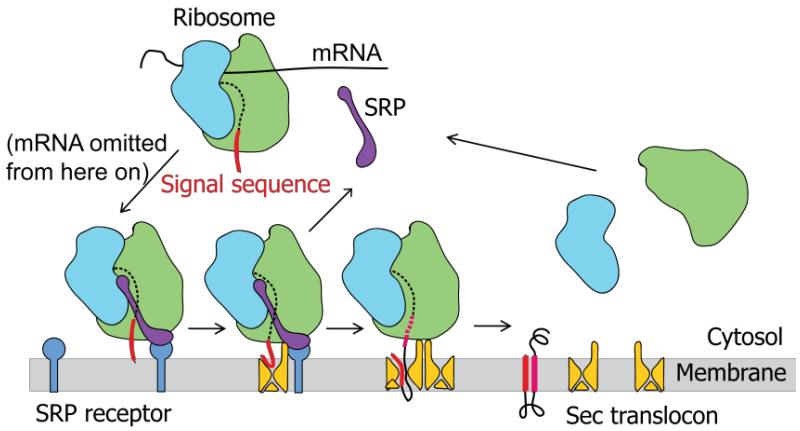


Figure 4. SRP-dependant co-translational insertion of an ER membrane protein. Adapted from (Rapoport, 2007).

1. **Targeting to the Sec translocon**. The SRP, a ribonucleoproteic cytosolic chaperone, will recognize and bind a highly hydrophobic stretch of the nascent polypeptide chain, the signal sequence (SS), as soon as it emerges from the ribosome. In eukaryotes, but not in prokaryotes, this arrests the chain elongation until the ribosome contacts the translocon. SRP then binds to its receptor (SR), an heterodimeric integral membrane protein complex, in the ER membrane (eukaryotes) or in the plasma membrane (prokaryotes), which catalyzes the transfer of the ribosome-nascent chain complex to the translocon (Nagai et al., 2003).
2. **Translocation/integration**. When bounds to the translocon, the exit tunnel in the ribosome is aligned with the translocon pore (Beckmann et al., 1997). After subsequent reiniziation of nascent chain elongation, the signal peptide reaches the translocon. If the nascent polypeptide chain is a secreated protein, the translocon facilitates protein translocation. In the case of integral membrane proteins, after reaching the translocon, the signal peptide either reorients, placing the N-terminal in the cytosol, giving the protein an Nin-orientation (type II), or it does not, resulting in an Nout-orientation (type I), see **Figure 5**, steps 3 and 2, respectively. Because type II signal peptide translocate their C-terminal flanking residues into the ER lumen, they must invert end-over-end 180º after exiting the ribosome in order to achieve their proper topology in the bilayer. Little is known about when and how such an inversion might take place in the context of the assembled ribosome translocon complex (RTC). The final topology of the signal peptide is the product of several factors that include (1) flanking charged residues, (2) hydrophobicity, (3) folding of the N-terminal domain, (4) attachment of N-linked carbohydrates, and (5) composition of membrane lipids (Bogdanov et al., 2002; Goder et al., 2004; Higy et al., 2004; Saurí et al., 2009).

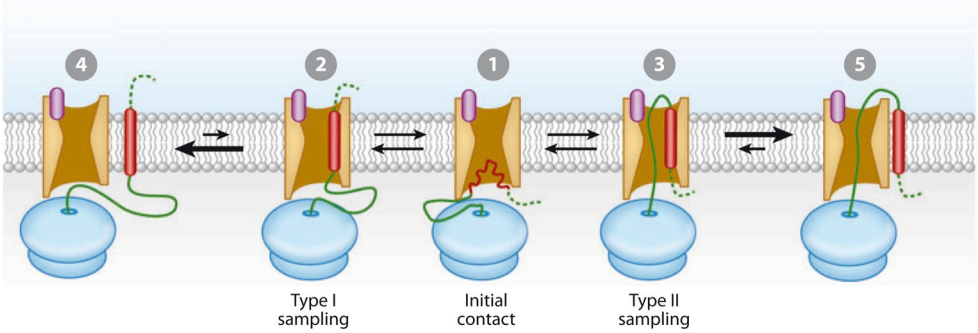


Figure 5. Models of transmembrane domain (TMD) insertion. (A) Generic model for TMD-translocon interactions. Upon its initial emergence from the ribosome, a signal peptide or a TMD (*red*) enters the cytosolic vestibule of the Sec61 channel (1). This initial metastable state allows the nascent chain to sample either of two orientations (2, 3). These states are envisioned to be interconvertible for a limited period of time, until the TMD laterally moves into the lipid bilayer in one or the other orientation (4, 5). Taken from (Shao and Hegde, 2011)

As translation and chain elongation proceeds, the nascent chain passes through the translocon channel. In this process, segments that are of sufficient length and hydrophobicity will enter into the lipid bilayer (Heinrich et al., 2000).

i.2.2. Translocon structure

The translocon complex is a marvelous transporter, since it is the only described channeling machine capable of permit passing of molecules through the membrane in two directions, across (perpendicular) and into (parallel/lateral) the membrane. Therefore, the co-translational pathway 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 (Rapoport, 2008). The gating capability of this complex in two directions (i.e. across the membrane and laterally into the lipid bilayer) differentiates and highlights it from the rest of the cellular channels.

In mammalian cells, this proteinaceous complex is composed of the Sec61 α-subunit, β-subunit and γ-subunit plus the translocating chain-associating membrane protein (TRAM) (Görlich et al., 1992). As translocon activity can be reproduced by *ab initio* reconstitution of these four membrane proteins in pure lipids (Görlich and Rapoport, 1993), these proteins constitute the core components of the mammalian translocon (Johnson and van Waes, 1999), see **Figure 6**.

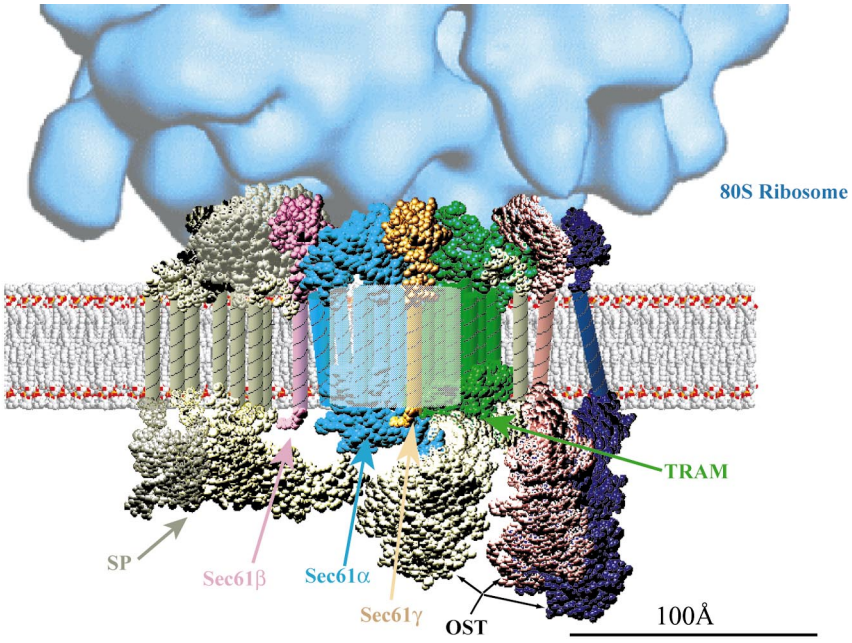


Figure 6. Protein packing in the mammalian translocon. (A) The components of the translocon are depicted approximately to scale in this cross section that is perpendicular to the plane of the membrane. TM segments are represented by a cylindrical volume with the dimensions of an average α-helix (12-Å diameter), whereas the cytoplasmic and lumenal domains of each protein are modeled using the dimensions of globular proteins or of portions of proteins with the same number of amino acids (the three-dimensional structures of ubiquitin, phospholipase, and bacteriorhodopsin were used as models for globular domains and α-helices). The shape of each domain is arbitrary and is shown merely to indicate the relative amounts of space occupied by translocon components on each side of the membrane. The relative positions of the polypeptides were chosen to be consistent with current models of a working translocon, so proteins known to interact with each other were placed in mutual proximity. Where no data regarding location are available, the placement was arbitrary. For clarity, only a single Sec61 heterotrimer is depicted. A ribosome is also shown to scale (From the surface reconstruction in Beckmann et al 1997) to indicate its size relative to the translocation machinery and the pore, depicted here as a 50 Å box with faded coloration. Taken from (Johnson and van Waes, 1999)

i.2.2.1. Sec61 complex

The eukaryotic Sec61 complex is a heterotrimeric membrane protein complex (Sec61 α, Sec61 β and Sec61 γ), called SecYEG in bacteria and archaeons. The α- and γ-subunit are highly conserved in all kingdoms, and are required for survival in both *Escherichia coli* and *Saccharomyces cerevisiae*. The β-subunit, on the other hand, is not required for survival, and does not have significant sequence homology between eukaryotes and eubacteria. The high-resolution structure of mammalian Sec61 translocon is not yet available. However, we have the homologous structures from *Methanococcus jannaschii* (van den Berg et al., 2004), *Thermus thermophilus* (Tsukazaki et al., 2008), *Thermotoga maritima* (Zimmer et al., 2008) and *Pyrococcus furiosus* (Egea and Stroud, 2010), 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 (Becker et al., 2009), and of the cryo-EM structure of SecYEG from *E. coli* with the mammalian Sec61 in a resting state (Ménétret et al., 2008), indicate a high degree of structural similarity between all Sec complexes.

i.2.2.2. The α-subunit

Sec61α is the main component of the protein-conducting channel of the translocon complex. This protein is a 53 kDa integral membrane protein, crossing the membrane 10 times, with both its N-terminus and C-terminus facing the cytosol (Wilkinson et al., 1996). Viewed from the top, the protein adopts a square shape that can be divided into two pseudosymmetric halves (**Figure 7**A), the N-terminal half containing TM segments 1–5 and the C-terminal half comprising TM segments 6–10 (red and blue TM segments in **Figure 7**, respectively). These two parts form an indentation in the centre through which the nascent chain passes, and, as mentioned above, is aligned with the ribosomal exit tunnel in a translocating translocon (Becker et al., 2009). From a lateral view, Sec61α has a rectangular contour and the channel within an hourglass shape (**Figure 8**A) (Rapoport et al., 2004).

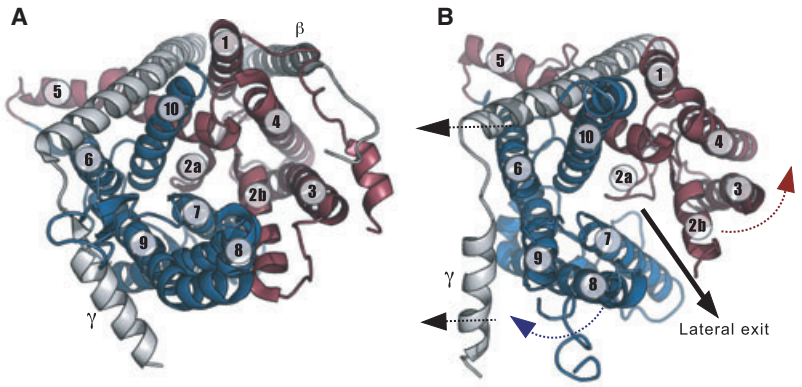


Figure 7. Translocon structure. Top view of the translocon structure. (A) Closed structure of the translocon from *M. jannaschii* (Protein Data Bank ID code: 1RHZ) (van den Berg et al., 2004). (B) Partially open structure of the translocon from *P. furiosus* (Protein Data Bank ID code: 3MP7) (Egea and Stroud, 2010). In both panels, all TM segments of Sec61α are colored (red and blue for each half; see text) except for the β-subunits and γ-subunits, which are shown in gray. All TM segments are numbered for easy comparison between the open and closed structures. The dotted arrows in (B) indicate the 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. Taken from (Martínez-Gil et al., 2011)

When it is in an inactive state, the cytoplasmic entry to the channel has a diameter of ~ 20–25 Å (van den Berg et al., 2004). Close to the middle of the membrane, the translocation pore reaches its narrowest point (5–8 Å), composed of a ring of bulky hydrophobic residues followed by a short helix (TM segment 2a) that blocks the channel pore (**Figure 8**). After this 'plug', the channel widens again towards the ER lumen.

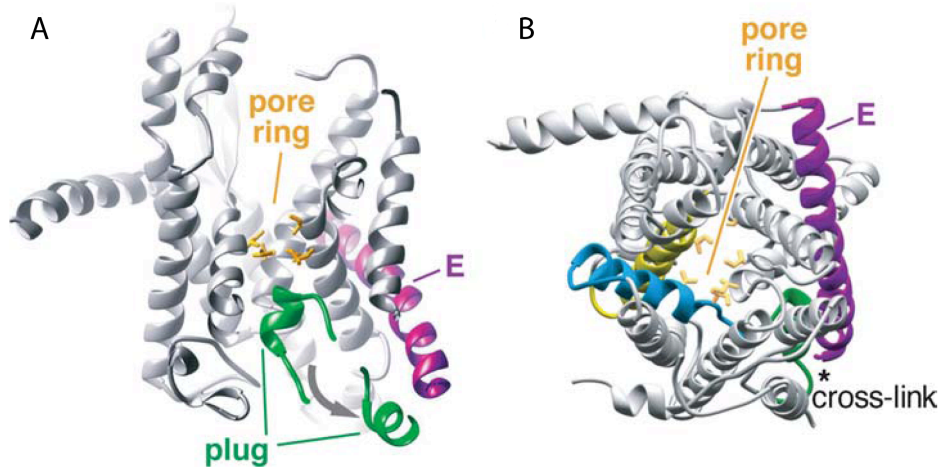


Figure 8. Plug movement would lead to opening of the SecY channel. (A) View from the side of the channel with the front half of the model cut away. The modeled movement of the plug toward the SecE subunit is indicated by an arrow. The side chains of residues in the pore ring are colored in gold. (B) Cytosolic view, with the plug modeled in its open position. TM2b and TM7 located at the front of the complex are shown in blue and yellow, respectively. The asterisk indicates the region where introduced cysteines result in cross-links between the plug and the TM segment of SecE (Harris & Silhavy 1999). Taken from (van den Berg et al., 2004)

Nevertheless, it has been reported that there is a significant increase in the pore diameter (Hamman et al., 1997), which is probably needed to accommodate the multiple TM segments of multispanning nascent chains that may leave the translocon in pairs or groups for membrane integration.

Even though a TM sequence is detected first by the ribosome, the translocon must still independently recognize the TM sequence to position it properly for its lateral partition into the bilayer. How does a TM sequence move laterally through the translocon and into the bilayer? Different stages of integration were examined by creating integration intermediates with nascent single-spanning membrane proteins of increasing length and incorporating a single photoreactive probe in the middle of the TM sequence of each to determine its environment (Do et al., 1996). Furthermore, site-directed photo cross-linking

studies revealed that TMDs consistently interact with Sec61α early in the insertion process, before or simultaneously with their interactions with lipids (Do et al., 1996; Heinrich et al., 2000; Martoglio et al., 1995; Mothes et al., 1997), which suggests that the Sec61 complex forms an interface between proteinaceous and lipid environments for inserting TMDs. This is consistent with the simplest case of IMP insertion, in which the translocon passively facilitates TMD partitioning between its aqueous interior and the surrounding hydrophobic bilayer. A mechanistic model for how lateral release of TMDs into the lipid bilayer occurs was suggested by high-resolution crystal structures of archaeal and bacterial homologs of the Sec61 complex (Tsukazaki et al., 2008; van den Berg et al., 2004)

After entering the translocon, the TM sequence was initially adjacent to both Sec61α and TRAM, but it was then moved sequentially to two different sites within the translocon from which only TRAM was photo-crosslinked. It is surprising that the TM sequence remained adjacent to TRAM until translation of the single-spanning membrane protein was terminated (Johnson and van Waes, 1999). These data suggest that the TM sequence moves through the translocon via a multistep pathway, by use of protein-protein interactions to regulate the location of the TM sequence in the translocon until the TM sequence is released into the bilayer after translation terminates, see **Figure 9**. If the latter model is correct, at least for some proteins, then what happens with the TM sequences in a multispanning membrane protein? Does the entry of a second TM sequence into the translocon cause the first to leave? The earliest models for polytopic IMP insertion, proposed more than 30 years ago, envisioned the successive integration of each TMD as each emerges from the ribosome (Johnson and van Waes, 1999). This sequential-insertion model is clearly feasible and necessitates that every TMD and its local sequence elements be sufficiently robust to drive both its translocon recognition and membrane insertion. Because TMDs in a polytopic IMP need to alternate orientations, each TMD should contain key topology determinants such as asymmetric flanking charges. This together with constraints imposed by the preceding TMD would facilitate proper biogenesis. This may well occur for some proteins, and it is feasible that sequential insertion is relatively common in bacterial systems where topological determinants are much more rigorously observed and the insertion machinery appears to be simpler.

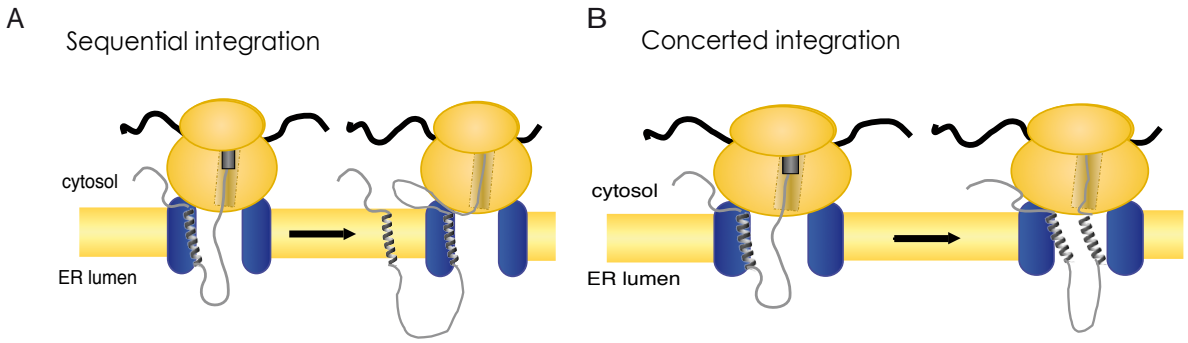


Figure 9. Models of polytopic membrane protein integration. A) Successive integration of each TM segment as it emerges from the ribosome. B) Both TM segments are in proximity to Sec61α, even when the nascent chain was long enough for the TM segments to exit the translocon individually, thereby indicating that they integrate together into the lipid bilayer as a helical hairpin.

As discussed above, eukaryotic polytopic IMPs are far more structurally diverse, and their TMDs are less stereotypic than are prokaryotic IMPs. This may be a consequence of their need to be more functionally dynamic with respect to diverse interactions and regulation, thereby imposing considerable sequence constraints that can clash with those needed for insertion. These same features may also be the reason why eukaryotic IMPs are poorly expressed in heterologous systems and difficult to crystallize relative to their prokaryotic counterparts. These and other considerations mean that poorly recognizable TMDs that in isolation are not able to insert must nonetheless be inserted in the context of the native IMP (Enquist et al., 2009). Several variations on this theme can be further envisaged. First, rather than being skipped entirely, candidate TMDs may be stored temporarily at or near the site of integration without necessarily committing to complete release into the membrane (Do et al., 1996; Sadlish et al., 2005; Saurí et al., 2007; 2005). Such TMDs could be held by accessory proteins or part of the Sec61 complex. Such provisional TMDs could at a later time, by the same mechanism outlined above, integrate or even reorient contingent on downstream events (Kauko et al., 2010). In more elaborate scenarios (building on essentially the same principles), multiple TMDs could be stored or skipped temporarily, only to reorient and integrate well after their initial encounter with the translocon. In the extreme case, reorientation and insertion of a subset of TMDs would occur completely posttranslationally in a process analogous to soluble protein folding (Kauko et al., 2010; Skach, 2009) .

i.2.2.3. The β-subunit

The β-subunit is a 9.9 kDa membrane protein that has a single TM domain located next to TM segments 1 and 4 of Sec61α (**Figure 7**A). Although this subunit is not essential either for translocation across the ER membrane or for insertion of TM segments into the lipid bilayer, it has been reported to kinetically facilitate cotranslational translocation (Kalies et al., 1998), and to interact with the SR heterodimer, probably facilitating recognition of unoccupied translocons by the RNC–SRP–SR complex (Jiang et al., 2008). The participation of Sec61β in the translocation process is also supported by its direct interaction with the nascent chain and the ribosome (Levy et al., 2001).

i.2.2.4. The γ-subunit

Sec61γ is the smallest component (7.7 kDa) of the Sec61 complex. It contains two helices connected by an extended loop (**Figure 7**). The first helical region, an amphipathic helix, sits parallel to the cytosolic side of the membrane and contacts the cytoplasmic side of the Sec61α C-terminal half. The second helix crosses the membrane diagonally, interacting with both N-terminal and C-terminal parts of Sec61α, and acts as a clamp that brings both halves of Sec61α together (van den Berg et al., 2004).

i.2.3. Translocation and insertion of a nascent chain

During cotranslational 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 subject of great controversy over the years. Initial cryo-EM studies indicated that three or four copies of the Sec61 complex could interact with the ribosome at the same time (Beckmann et al., 1997). However, biochemical studies and the structures that have recently become available (see above) strongly suggest that only one copy of the Sec61–SecY complex is required for translocation (Becker et al., 2009; Cannon et al., 2005; Egea and Stroud, 2010; Ménétret et al., 2008; Park and Rapoport, 2012; van den Berg et al., 2004; Yahr and Wickner, 2000). Biochemical analysis of Sec61 point mutants, and the cryo-EM reconstructions of the ribosome–translocon pair, indicate that the loops between TM segments 6 / 7 and 8 / 9 of the Sec61α-subunit (or its homologs) are involved in this association (Becker et al., 2009;

Ménétret et al., 2007). In fact, point mutations within those loops of *E. coli* SecY are known to affect the ribosome–SecY interaction (Ménétret et al., 2007). However, similar changes in the equivalent loop of the yeast translocon did not affect binding to the ribosome (Becker et al., 2009). All of this indicates that, despite small differences, the ribosome–Sec junction is well conserved among species, and that probably only one translocon complex is bound to a translating ribosome.

Although many details remain unknown, significant insights into the mechanism of membrane insertion have come from structural studies. The process starts with the engagement between the translocon complex and its cytosolic partner (i.e. the ribosome in the cotranslational pathway). Either this contact or the presence of the signal sequence (SS) in the nascent chain or both, triggers the widening of the cytosolic side of the channel (Tsukazaki et al., 2008), including the hydrophobic ring, which increases from ~5 to ~ 14 Å (**Figure 7**B)(Egea and Stroud, 2010). 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 interface between TM segments 2b and 7⁄8), 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 (**Figure 10**), 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 the phospholipids through the already open lateral crack (Plath et al., 1998). As the elongation of the nascent chain continues, two rearrangements occur in Sec61α. First, the plug is 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 half move apart from each other (**Figure 7**B), creating a lateral gate across the entire channel, which exposes the nascent polypeptide to the core of the membrane (Egea and Stroud, 2010; Martoglio et al., 1995). The polypeptide sequence within the translocon can then partition into the lipids if it is hydrophobic enough, as the SS would do, or continue through the translocon into the ER lumen for hydrophilic translocated regions. The structural changes in the α-subunit are accompanied by a dramatic shift (**Figure 7**B) in the location of the N-terminal helix of Sec61γ⁄SecE, which

releases the clamp over Sec61α. Nevertheless, the opening of the lateral gate is not required to accommodate a translocating peptide within the channel (Becker et al., 2009). 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, see **Figure 10**. 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, while the opening/closing of the lateral gate exposes hydrophobic segments to the lipid bilayer, allowing their partition into the membrane.

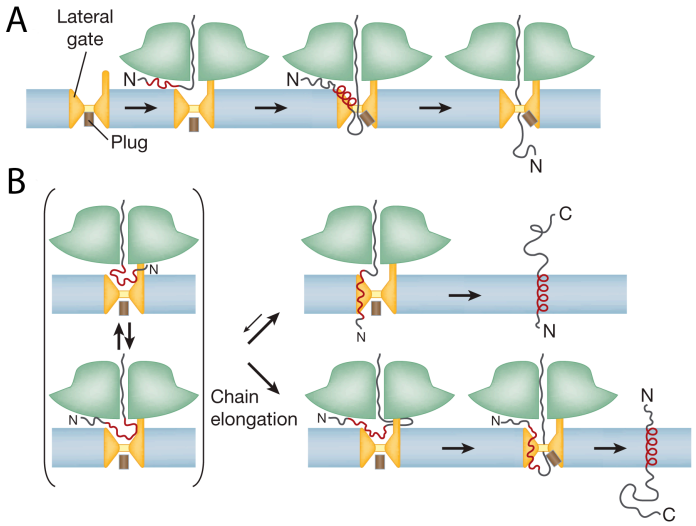


Figure 10. Different stages of translocation. (A) Translocation of a secretory protein. The red line indicates the hydrophobic region of a signal sequence. Depicted is the co-translational mode of translocation, but similar schemes can be envisioned for the other modes. For simplicity, only the translocating Sec61/SecY copy is shown. (B) Insertion of membrane proteins. When a hydrophobic TM sequence (in red) has emerged from the ribosome, it can bind reversibly in several conformations. If the hydrophobic sequence is long and the N terminus is not retained in the cytosol, it can flip across the membrane (upper panel). If the N terminus is retained in the cytosol and the polypeptide chain is further elongated by the translating ribosome (indicated by the loop between the ribosome and channel), the C terminus can translocate across the membrane (lower panel). Taken from (Rapoport, 2007)

i.2.4. Translocon associated proteins

i.2.4.1. TRAM

The translocating chain-associated membrane protein (TRAM), was initially identified by crosslinking methods in reconstituted proteoliposomes (Görlich et al., 1992). Although it is 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 eight TM segments and both the N-terminus and C-terminus facing the cytosol (Tamborero et al., 2011). The role of TRAM in the translocation of secretory proteins is restricted to the insertion of the SS into the membrane (Voigt et al., 1996), where TRAM has been found to be required for the insertion of SSs with either short hydrophobic regions or with low overall hydrophobicity. Regarding the insertion of TM segments, TRAM has also been reported to cross-link with a wide variety of TM segments (Do et al., 1996; Heinrich and Rapoport, 2003; Martínez-Gil et al., 2010; McCormick et al., 2003; Saurí et al., 2007), some of them containing charged residues (Cross and High, 2009; Heinrich et al., 2000; Meacock et al., 2002). These observations, together with the fact that TRAM itself contains an unusually 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 at the initial stages of membrane integration (Tamborero et al., 2011).

i.2.4.2. Oligosaccharyl transferase

Eukaryotic cells are stringently regulated and controlled through the compartmentalization of cellular processes, including cell-cycle control (nucleus), energy production (mitochondria or chloroplast), protein processing (endoplasmic reticulum (ER) and Golgi apparatus) and waste removal (lysosomes). Protein processing in the secretory pathway is essential for proteins destined to be released from the cell or integrated into cellular membranes. More than 70% of all proteins that are processed by these organelles are glycosylated (Gavel and Heijne, 1990). The complexity of the oligosaccharides found decorating glycoproteins is astonishing when compared with other cellular constituents, such as RNA and protein, because of the variety of chemical linkages that can be formed to a single saccharide building block. In keeping with this molecular diversity, the population of carbohydrate structures displayed by any one protein is not homogenous, thus adding to the complexity of glycoprotein structures.

In *N*-linked glycosylation, oligosaccharyl transferase (OST, EC 2.4.1.119; also known as: dolichyl-diphos-phooligosaccharide-protein-glycosyltransferase) transfers a tetradecasaccharide (GlcNAc2Man9Glc3) from a dolichol pyrophosphate donor to selected asparagine side chains (in an Asn-Xxx-Ser/Thr consensus sequence, where Xxx is any

amino acid except proline) within nascent polypeptides in the lumen of the ER. During translocation, the nascent protein is directed towards OST. Nascent proteins have been shown *in vitro*, to be glycosylated only once the Asn-Xaa-Thr/Ser consensus sequence for *N*-linked glycosylation has cleared 30–40 Å from the face of the ER membrane (Nilsson and Heijne, 1993a).

The OST is a 700 kDa multimeric protein complex wherein many of the proteins are highly homologous, from protozoan to mammalian species. The most studied OST complex is that derived from the budding yeast *S. cerevisiae.* This complex includes nine different membrane-bound subunits (**Nlt1p/Ost1p, Ost2p**, Ost3p, Ost4p, Ost5p, Ost6p, **Stt3p**, **Swp1p and Wbp1p**), five (in bold) of them have been determined to be essential for cell viability (Dempski and Imperiali, 2002).

i.2.4.3. Signal Peptidase

In eukaryotes, proteins that are targeted to the ER membrane are preceded by signal peptides that target the protein either cotranslationally or post-translationally to the Sec61 translocation machinery. The ER signal peptides have features similar to those of their bacterial counterparts. Signal peptides are cleaved from the exported protein after translated into the ER lumen (or across the plasma membrane in prokaryotes), by the signal peptidase (SP) complex. Signal peptides that sort proteins to different locations within the eukaryotic cell have to be distinct because these cells contain many different membranous and aqueous compartments. Proteins that are targeted to the ER often contain cleavable signal sequences. The hydrophobic character is important for function as a translocation signal (Cioffi et al., 1989). Hence, a certain threshold is needed to be reached for efficient translocation. ER signal peptides have a higher content of leucine residues relative to bacterial signal peptides (Nielsen et al., 1996). The signal recognition particle (SRP) binds to cleavable signal peptides after they emerge from the ribosome (Ng et al., 1996; Zheng and Gierasch, 1996), and targets the nascent protein to the ER membrane. After translocation of the protein to the ER lumen, the exported protein is processed by the SP.

While many proteins are targeted by the signal peptide through the SRP route, some proteins are targeted by an SRP-independent route, which involves the Sec62p-Sec63p complex at least in yeast.

i.2.4.4. Other Translocon-associated proteins

Some other membrane proteins [i.e. translocon-associated protein (TRAP), PAT-10, RAMP4 and 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 to form part of the translocon core complex.

TRAP, translocon-associated protein (150 kDa), is a tetrameric protein complex (α, β, γ and δ) of integral membrane proteins (Hartmann et al., 1993). It is associated with ribosome–Sec61 complexes with a 1:1 stoichiometry (Ménétret et al., 2008). It has been proposed that TRAP facilitates the initiation of protein translocation (Fons et al., 2003), although the details of the mechanism remain unknown. The α, β, and δ subunits are single spanning membrane proteins, whereas the γ subunit has been proposed that crosses the membrane four times (Hartmann et al., 1993).

PAT-10 a 10 kDa protein, described as a translocon-associated protein during a search for Sec61 partners for opsin nascent chain insertion as a bait (Meacock et al., 2002). It is a membrane protein that crosslinks with some of the opsin TM segments (Ismail et al., 2008). 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 each particular TM segment within the opsin nascent chain.

RAMP4 is a small (66-residue long) tail-anchored membrane protein implicated in promoting correct integration/folding of integral membrane proteins by facilitating subsequent glycosylation (Yamaguchi et al., 1999). 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 that triggers this recruitment (Pool, 2009).

Another protein that has been reported to interact with the translocon complex is **BAP31**. This multispanning integral membrane protein participates in the identification of misfolded proteins at the ER and their retrotranslocation to the cytoplasm. The finding that BAP31 interacts with both Sec61β and TRAM (Wang et al., 2008) suggests a role of the translocon in membrane protein quality control.

The increasing number of interacting partners of the translocon also indicates that a variety of functions of the channel may be performed in association with different cellular components. Indeed, the Sec61 complex might be merely the common player in a wide variety of transient complexes, each one performing different but related functions.

i.2.5. Post-translational translocation

The Sec translocon is also involved in post-translational translocation of proteins. After protein synthesis, aggregation is prevented by cytosolic chaperones binding to the fully synthesized protein containing signal sequences or TM regions, before it is transferred to the translocation machinery. In eubacteria like *E. coli*, the translocating protein is pushed through the translocon pore in an ATP-demanding process by the cytosolic SecA protein instead of the ribosome. A complex of SecD, SecF and YajC also participates in the process (Erlandson et al., 2008; Zimmer et al., 2008). SecA is also needed for the translocation of the extracellular domains of several membrane proteins utilizing the co-translational pathway (**Figure 11**B). In yeast, and most likely, in all eukaryotes, the translocating protein moves instead through the pore by Brownian motion after the signal sequence binds to the translocon. The luminal chaperone and ATPase BiP (Immunoglobulin heavy chain-binding protein homolog) (78kDa) binds the protein and prevents movement back into the cytosol. This process also requires the assistance of the additional transmembrane Sec62/63 complex, see **Figure 11**A.

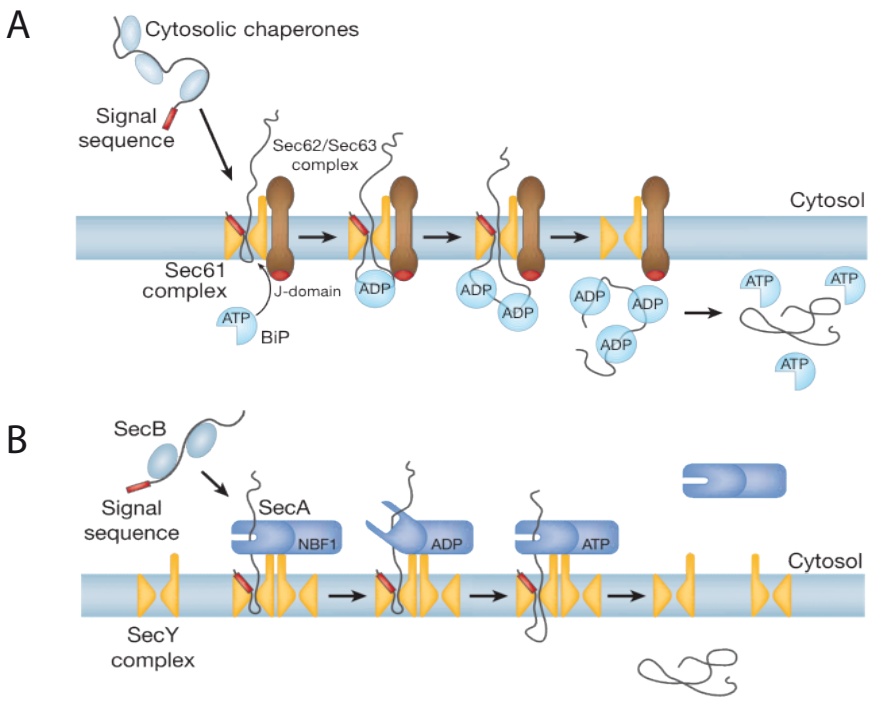


Figure 11. Post-translational translocation. (A) Model of post-translational translocation in eukaryotes. It is possible that oligomers of the Sec61 complex mediate translocation, similar to the situation with the other modes of translocation. (B) Model of post-translational translocation in bacteria. Taken from (Rapoport, 2007)

Translational termination of C-terminal tail anchored (TA) proteins make them poor SRP substrates, necessitating for their recogintion, targeting, and insertion a purely posttranslational mechanism. Recent observations in yeast, suggest that the GET pathway in yeast plays an essential role in TA biogenesis **(Figure 12)**. Without Get1 and/or Get2, Get3 is cytosolic. *In vitro* reconstitutions showed that Get1/2 recruit Get3 to the ER membrane in an ATP-dependent manner (Schuldiner et al., 2008). Thus, Get1 and Get2 are thought to form an ER-localized complex that serves as a receptor for Get3-mediated TA-protein targeting.

Biochemical studies revealed that the mammalian Get3 homolog, TRC40/Asna1, binds the TA protein, Sec61β, and facilitates its post-translational insertion into ER membranes.

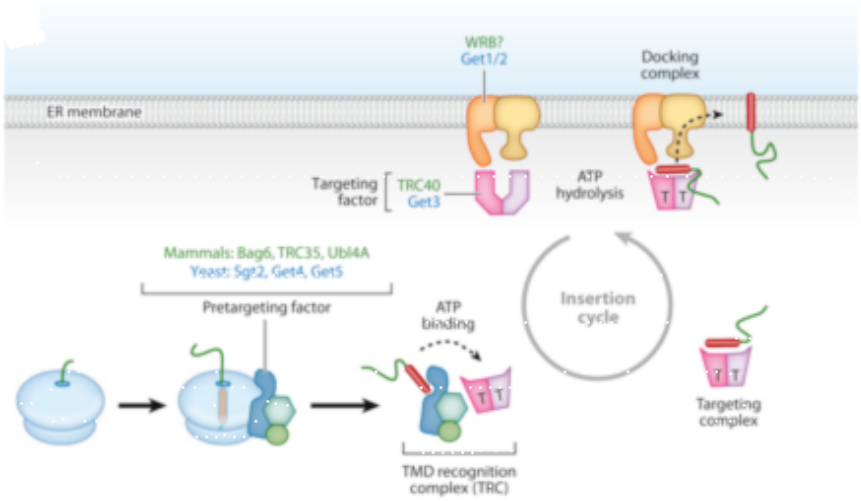


Figure 12. Tail-anchored (TA) protein insertion. Schematic model of the known components and steps mediating posttranslational insertion of a TA protein. When the TM domain of a TA protein is synthesized, it favors recruitment of a pretargeting factor to the ribosomal surface. This is composed of Bag6, TRC35, and Ubl4A in mammals. The analogous complex in yeast is formed by Sgt2, Get4, and Get5 as well as other chaperones. Its location near the ribosome would favor capture of the TA protein upon its release. The pre-targeting factor together with the targeting factor (TRC40 in mammals, Get3 in yeast) (*pink*) form the TRC. This is thought to be a transient complex that facilitates sorting, recognition, and loading of the TA protein onto the targeting factor. The targeting factor is an ATPase, and its substrate-bound form is thought to be ATP-bound (indicated by a T). This is delivered to the ER membrane via a receptor composed of Get1 and Get2 in yeast (a mammalian homolog of Get1 may be WRB). The docking complex of Get1-2-3 somehow facilitates substrate release and insertion in a step that depends on ATP hydrolysis by Get3. The now-vacant Get3 (which is in a different open conformation) is recycled to the cytosol to complete the insertion cycle. Taken from (Shao et al, 2011)

i.2.6. Peroxisomal proteins and integration into the ER

Although peroxisomes appear to be simple organelles, their formation and maintenance pose unique challenges for the eukaryotic cell. Peroxisomes are, with mitochondria, the exclusive site for fatty acid breakdown and are essential for growth on oleate as the sole carbon source. However, on glucose, peroxisomes are dispensable.

Peroxisomes are unusual in this respect because their biogenesis requires both these assembly lines: (1) the ER provides lipids and peroxisomal membrane proteins (PMPs) and yields a peroxisomal precompartment (Hoepfner et al., 2005; Kragt et al., 2005; Tam et al., 2005; van der Zand et al., 2010) and (2) the cytosol provides the matrix proteins, which are imported via the peroxisomal translocon (Agne et al., 2003). Together these processes define the beginning and end of the peroxisomal biogenesis pathway

In early electron microscopy (EM) images, peroxisomes were often observed close to the endoplasmic reticulum (ER), and occasionally ER-peroxisome membrane continuities were claimed to exist, sparking the idea that the ER was involved in the biogenesis of

peroxisomes (Novikoff and Novikoff, 1972). Recently, have appeared strong molecular evidences supporting the hypothesis that the ER is indeed involved in peroxisome biogenesis (Agrawal et al., 2011; Hoepfner et al., 2005; Kragt et al., 2005; Lam et al., 2011; van der Zand et al., 2010; 2012).

Unity in nomenclature was agreed upon and the genes involved in proper peroxisome assembly were called PEX genes (Distel et al., 1996), which encode peroxins. The most interesting of these PEX genes coded for peroxisomal membrane proteins (PMPs), some of which turned out to be involved in organelle biogenesis. Hence, membrane assembly and maintenance requires three of these gene products (PEX3, PEX16 and PEX19) and may occur without the import of the matrix (lumen) enzymes.

The Δpex3 and Δpex19 deletion mutants, for instance, represent extreme cases: in the absence of these genes no trace of residual peroxisomes is left (Hettema et al., 2000). When wild-type *pex3* or *pex19* genes are introduced in Δpex3 or Δpex19 deletion mutants, respectively, the organelles reappeared despite many generations of growth in the absence of peroxisomes (Subramani, 1998). The appearance of Pex3 protein (PEX3) in the ER is not species specific but is a generally shared property in fungi and mammals (Kragt et al., 2005; Tam et al., 2005). It has been showed that PEX3 first appeared in the perinuclear ER and subsequently assembled in a punctate fluorescent dot coinciding with the ER but later disconnected from the ER (Hoepfner et al., 2005). At this stage, a fluorescent reporter for matrix protein import started to colocalize with PEX3, signaling the onset of the formation of mature, metabolically active peroxisomes. This whole process is accomplished in a couple of hours.

This trafficking route is representative of PMPs in general. At least 15 additional PMPs differing in function and membrane topology first inserted into the ER membrane before appearing in peroxisomes (van der Zand et al., 2010). We can extrapolate from this number and consider the ER-to-peroxisome trafficking route as exemplary for all other PMPs. This same group of PMPs was studied not only in Δpex3 and Δpex19 deletion mutants, which restored their peroxisome population upon functional complementation with PEX3 or PEX19, but also in wild-type cells already containing peroxisomes. These

findings demonstrate that the ER's contribution in peroxisome formation is a general operating principle.

Most PMPs enter the ER via the Sec61 complex. This was shown *in vivo* by depletion assays, in which an essential component of the Sec61 translocon becomes limiting in time. The reduced protein-import capacity affected secretory proteins and PMPs to the same extent (van der Zand et al., 2010). *In vitro* experiments corroborated these observations, in reticulocyte lysate with yeast or dog pancreas microsomes, in which newly synthesized, radiolabeled PEX3 entered the ER membrane in a form resistant to extraction with sodium carbonate (Thoms et al., 2012).

As a prerequisite to understand these early steps in peroxisomal biogenesis, it is essential to ascertain how peroxins are inserted into the ER membrane. In yeast, two pathways have been identified as being involved: a small group of tail-anchored PMPs are most likely post-translationally inserted via the GET3 pathway (Schuldiner et al., 2008; van der Zand et al., 2010), while others appear to be inserted through the yeast Sec61 translocon (Thoms et al., 2012; van der Zand et al., 2010) that serves as the primary ER entry point for secretory and integral membrane proteins. However, these previous experiments did not reveal any mechanistic details about how the yeast Sec61 translocon facilitates PMP insertion into the ER bilayer. As is the case for most of the soluble proteins entering the Sec61 channel, proteolytic cleavage offers no clue. PMPs enter and leave the ER without evidence of processing. A first step in answering this question was taken using PEX3. In PEX3, a conserved N-terminal signal comprising the TM domain and its preceding positively charged amino acids was identified as important for Sec61-mediated import (van der Zand et al., 2010).

Some authors concluded that PMP cargo is selected in the ER and leaves the ER in separate groups via two different vesicular carriers (Tabak et al., 2012). Upon heterotypic fusion of these carriers, components of the translocon arrive in one organelle, allowing their association in a functional complex. Filling the matrix with enzymes finalizes the formation of metabolically active peroxisomes (**Figure 13**).

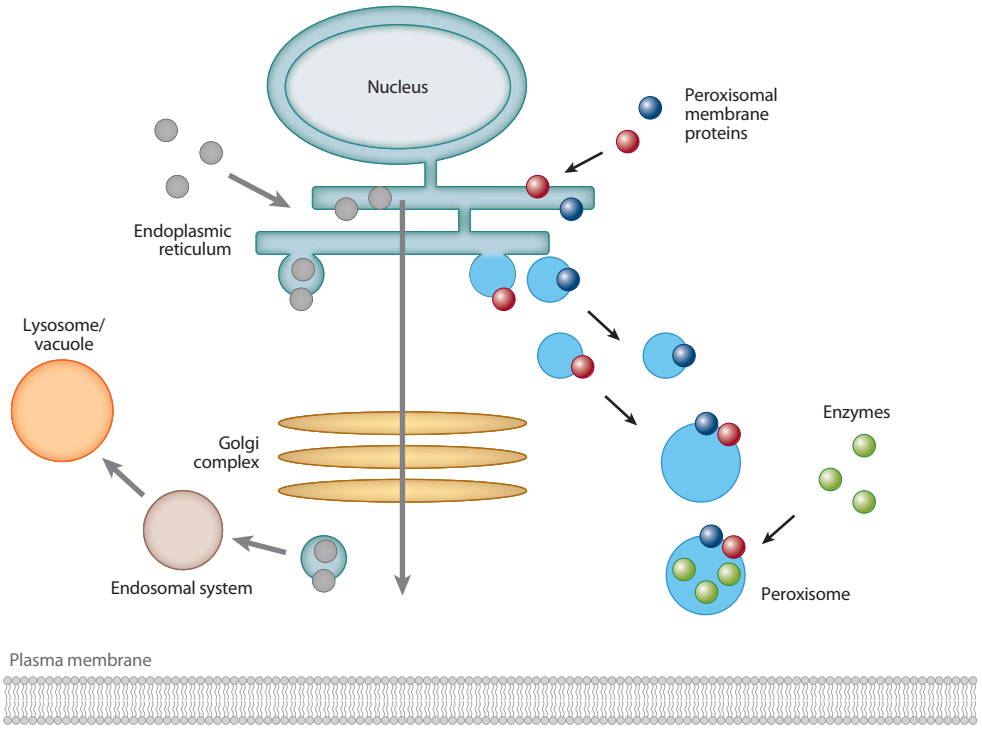


Figure 13. Vesicular flow transporting peroxisomal membrane proteins (PMPs) from the endoplasmic reticulum (ER) on the way to the formation of new peroxisomes. Secretory proteins and PMPs insert into the ER. Within the ER membrane, PMPs assemble into various subcomplexes (*blue* and *red circles*). These complexes are then recruited to separate ER exit sites. Budding from the ER results in the formation of different vesicle pools characterized by their unique PMP cargo. Preperoxisomal vesicles heterotypically fuse, leading to the assembly of the really interesting new gene finger (*blue*) and docking (*red*) subcomplexes into a full, functional peroxisomal translocon. This assembly process (van der Zand et al., 2012) is required for the subsequent import of cytosolic peroxisomal targeting signal containing enzymes (*green circles*) into the peroxisomal matrix and completes the maturation of a peroxisome into a metabolically active organelle. Proteins destined for the secretory pathway (*gray circles*) leave the ER via the Golgi complex to the plasma membrane or via the endosomal system to the lysosome/vacuole. Taken from (Tabak et al., 2012)

I.3. Structure and topology of α-helical membrane proteins

Membrane proteins can be described by their topology, what segments are TM and extramembrane; cytosolic or extracytosolic (lumenal, periplasmic, etc.)

Single-spanning membrane proteins may assume a final topology with a cytoplasmic N- and an exoplasmic C-terminus (N_{cyt}/C_{exo}) or with the opposite orientation (N_{exo}/C_{cyt}). However, if the mechanism of insertion is taken into consideration, four major types of single-spanning membrane proteins can be distinguished, as is summarized in **Figure 14**. Type I membrane proteins are initially targeted to the ER by an N-terminal, cleavable

signal sequence, a hydrophobic stretch of typically 7-15 predominantly apolar residues, and then anchored in the membrane by a subsequent stop-transfer sequence, a segment of 20 hydrophobic residues that halts further translocation of the polypeptide and acts as a TM anchor. In type II membrane proteins, a signal-anchor sequence is responsible for both insertion and anchoring. Signal-anchor sequences are generally longer than cleaved signals (18-25 mostly apolar amino acids), since they span the lipid bilayer as a TM helix. They lack a signal peptidase cleavage site and they can be positioned internally within the polypeptide chain. However, like cleaved signals, they induce the translocation of their C-terminal end across the membrane. The opposite is the case for reverse signal-anchors of type III proteins, which translocate their N-terminal end across the membrane. These three types of membrane proteins are all inserted by the same machinery involving SRP, SRP receptor and the Sec61 translocon. With respect to topogenesis, there are two basic types of signals translocating either their C-terminus (cleaved signals and signal-anchors) or their N-terminus (reverse signal-anchors).

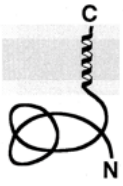
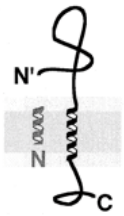
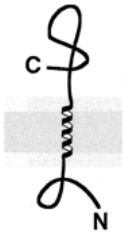
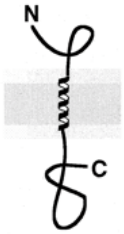
Signal type:	C-Terminus-translocating signals			N-Terminus-translocating
Topogenic determinants:	C-terminal signal	Cleaved signal + stop transfer sequence	Signal-anchor	Reverse signal-anchor
Machinery:	Get/TRC pathway	SRP/SR/Sec61 + signal peptidase	SRP/SR/Sec61	SRP/SR/Sec61
exoplasmic / Final topology: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

Figure 14. Topogenic determinants of single-spanning membrane proteins. Adapted from (Goder and Spiess, 2001)

In addition, there is also a class of proteins predominantly exposed to the cytosol and anchored to the membrane by a very C-terminal hydrophobic signal sequence. Examples are cytochrome b5, the SNARE proteins like synaptobrevin and several Bcl-2 family proteins. Insertion of these proteins is necessarily post-translational as mentioned above, since the signal emerges from the ribosome only after translation has reached the stop

codon. Accordingly, targeting and insertion for these C-tail anchoring proteins (TA) was found to be independent of SRP and the Sec61 complex (Shao and Hegde, 2011).

In an aqueous environment like the cytoplasm, the hydrophobic effect will cause hydrophilic residues and the polar peptide backbone to be hydrated by water molecules while hydrophobic parts are hidden in the protein interior. In the hydrophobic core of the membrane the situation is the opposite and the presence of hydrophilic residues and the backbone will be energetically unfavorable. Integral membrane proteins minimize the cost of harboring the polar polypeptide backbone within the membrane by engaging its polar groups in hydrogen bonds as early mentioned in this thesis, the two major structural classes of membrane proteins present two different ways of doing this. (1) β-barrel membrane proteins form cylinders of amphipathic β-strands and (2) α-helical membrane proteins span the membrane with one or more hydrophobic α-helices (**Figure 1**).

i.3.1. The primary structure of a TM helix is adapted to its surroundings

With respect to the different milieus a TM helix traverses, there is a statistical difference in the amino acids residues distribution in different regions along helix, see **Figure 15**. The majority of the side chains found within the core of the membrane are hydrophobic (Granseth et al., 2005; Hessa et al., 2007; Krogh et al., 2001). However, as more 3D-structures are obtained, many examples of charged residues and sequences causing irregular secondary structure effects within the hydrophobic core have been observed (Kauko et al., 2010).

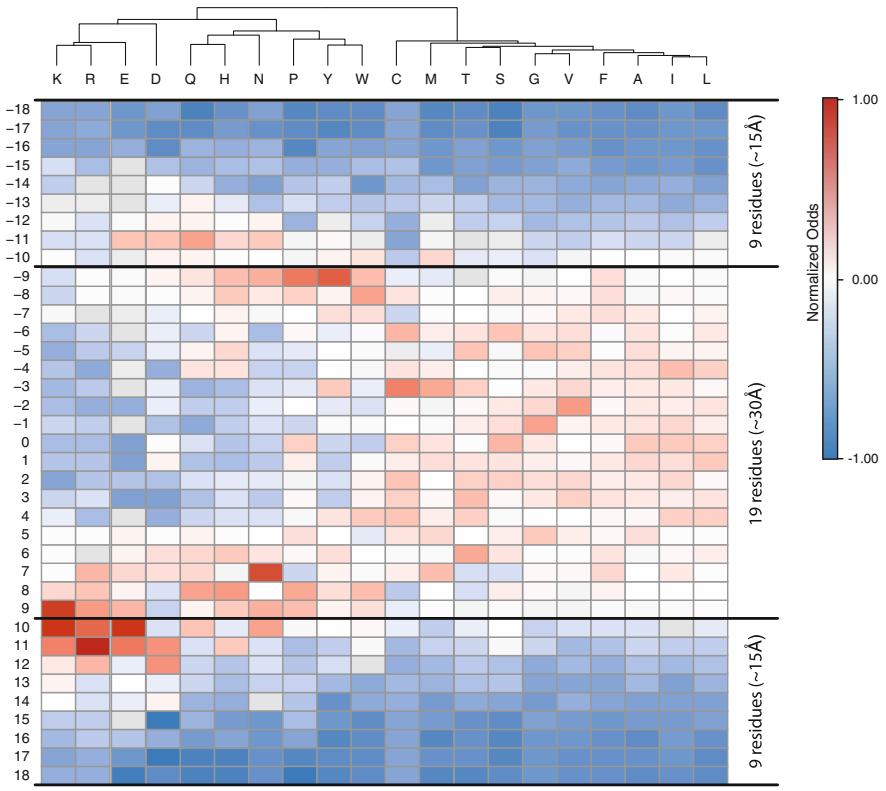


Figure 15. Amino acid type and position distribution in TM helices. Each amino acid type and its positioning in the TM helix is represented by its position-normalized Odds (that is, for each column the Odds are normalized to an average of zero and a standard deviation of unity). The amino acids are clustered on the basis of their positional normalized Odds within the helices. Positively labeled positions indicate the cytoplasmic side of the membrane and its flanking region whereas negatively labeled positions are indicative of extra-cytoplasmic regions. Taken from (Baeza-Delgado et al., 2013)

i.3.1.1. Charged and polar residues are better tolerated at the helix termini due to snorkeling

Charged or polar amino acids are energetically better tolerated towards the helix termini (Freites et al., 2005; Hessa et al., 2007). Their side chains will "snorkel", orienting the polar groups so that they approach the interfacial, aqueous regions (Chamberlain et al., 2004; Strandberg and Killian, 2003). This allows them to pull hydrating water molecules into the hydrocarbon part of the bilayer and create polar microenvironments for themselves, as seen in simulations (Johansson and Lindahl, 2006). Obviously a long side chain is advantageous for snorkeling (Jaud et al., 2009; Johansson and Lindahl, 2006) while shorter side chains would induce helix tilting, see **Figure 16**. However, not all polar or charged residues in TM segments are found in positions allowing snorkeling.

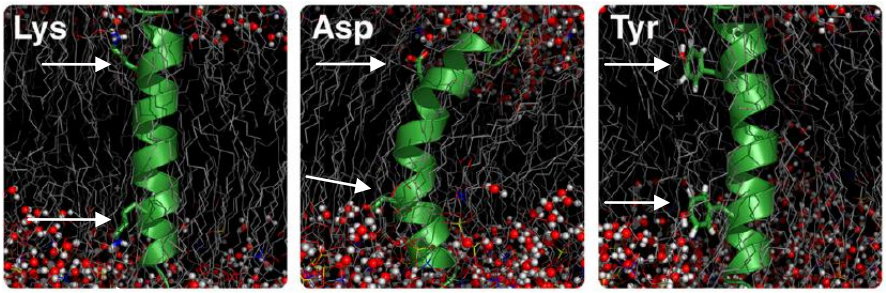


Figure 16. Snorkeling of side chains effects on helix tilting. Simulated model systems with the denoted amino acid placed symmetrically (indicated by arrows) with an offset of 5 positions from the center of the hydrocarbon core. Note the bend induced by the shorter Asp side chains. Adapted from (Johansson and Lindahl, 2006).

i.3.2. Forces and motifs implied in the folding of helical membrane proteins

The native fold of a membrane protein is mostly directed by interactions between its TM helices. The current understanding of the factors stabilizing a native fold is that van der Waals forces and hydrogen bonds dominate the process, aided by the occasional salt bridge and aromatic interactions. Also, the lateral pressure in the lipid bilayer, as well as the thickness of the bilayer, affects protein stability. Following, the intramolecular forces will be introduced.

i.3.2.1. Van der Waals forces

Van der Waals forces occur between all atoms and molecules. Non-polar molecules can become temporary dipoles if their total electron density gets unequally distributed as the electrons move about. If close enough, this dipole can induce a temporary dipole in the electron density of a neighboring non-polar molecule, allowing electrostatic interaction. The interaction gets stronger the closer the molecules are. The information on how tight TM helices pack in the membrane is ambiguous (Adamian and Liang, 2001; Eilers et al., 2000; Gerstein and Chothia, 1999; Hildebrand et al., 2005), but the residues in TM regions do tend to be buried more often than residues in soluble proteins or extramembrane regions (Oberai et al., 2009), allowing a higher number of van der Waals interactions if not necessarily stronger ones.

i.3.2.2. Hydrogen bonds

A hydrogen bond is the interaction between an atom with a non-bonded pair of electrons (O, N, etc.) and a hydrogen atom covalently bonded to a strongly electronegative atom (N, F, O). The electronegative atom attracts the electron density of the small hydrogen atom, decentralizing it and leaving the hydrogen with a substantial positive partial charge. Carbon can also take the place of the electronegative atom, resulting in a weaker hydrogen bond. In membrane proteins, the TM helix backbones can be involved in such weak hydrogen bonds, $C_{\alpha}H\cdots O$. It is likely to be substantial for helices at close distances. Strong hydrogen bonds involving Asn, Asp, Gln and Glu have been shown to drive helix oligomerization (Choma et al., 2000; Gratkowski et al., 2001; Zhou et al., 2000; 2001). The strength of any type of hydrogen bond depends on the distance, the chemistry and relative arrangement of the involved atoms, and the nature of the surrounding milieu, which is highly hydrophobic at the membrane core

i.3.2.3. Ion pairing/salt bridges

Salt bridges are electrostatic interactions between ions of opposite charge. Within the hydrophobic milieu of the lipid bilayer, the dielectric constant is very low and the interaction of a salt bridge is therefore stronger. Charged residues (protonated Lys, Arg, His and deprotonated Asp, Glu) are not common in TM protein segments (Baeza-Delgado et al., 2013), but they are present. All together, these residues constitute about 6.6% of the residues within TM helices.

i.3.2.4. Sequence motifs allowing tighter packing of helices, thereby increasing stabilization

A number of motifs involving small side chains allow tight packing of TM helices. For this packing, small residues are spaced in such a way that their side chains end up on the same side of the helix, either i, i+4 or i, i+7, creating a surface that allows close proximity. Statistics show that glycine zippers like GxxxG, (G, A, S)xxxGxxxG and GxxxGxxx(G, S, T) occur in more than 10 percent of all known MP structures (Kim et al., 2005). These motifs usually form right-handed crossings of the helices. The GxxxG motif has been extensively studied (Kim et al., 2005; Lemmon et al., 1992; MacKenzie et al., 1997; Mingarro et al., 1996) and can readily, depending on sequence context, induce dimerization

of TM helices (Doura et al., 2004; Langosch et al., 1996; Orzáez et al., 2000). However, the majority of helix-helix interactin interfaces seem to be formed by small residues in a i, i+7 pattern, resulting in left-handed crossing at small angles and knobs-into-holes packing, where side chains i and i+3 are interdigitated between four side chains of the other TM helix (Psachoulia et al., 2008). Tightly packed helices are stabilized by stronger van der Waals interactions due to closer proximity of the interacting helices. Many helix-helix interfaces are also stabilized by hydrogen bonds, both "classical" bonds and the weaker bonds involving the peptide backbone, Cα-H···O (Adamian and Liang, 2001). It has been also shown that the residues participating in hydrogen bonds are more conserved than other buried residues (Hildebrand et al., 2005), and that the packing preferences differ between channels and transporters and other membrane proteins, mostly because these class of membrane proteins are enriched in different amino acids. In channels and transporters, the helix-helix interfaces are not as tightly packed since they have to leave room for the pass of crossing molecules. Also, hydrogen bonds occur more frequently and are often found in proximity to water-filled cavities providing alternative bonding partners, allowing weaker bonds suitable for the structural rearrangements needed for functionality.

i.3.2.5. Lipid-protein interactions affecting the folded protein

The lipid composition of the bilayer affects the conformation of membrane proteins. Although protein sequence appears to be the primary determinant of final organization, the topology of several twelve-TM spanning secondary transporters of *E. coli* is dramatically influenced by the membrane lipid composition. The N-terminal six-TM helical bundle of *E. coli* Lactose permease (Bogdanov et al., 2002) (see **Figure 17**) the N-terminal two-TM hairpins of phenylalanine permease (PheP) (Zhang et al., 2003) and γ-aminobutyrate permease (GabP) (Zhang et al., 2005) are inverted with respect to the membrane bilayer when assembled in membranes lacking the major lipid phosphatidylethanolamine (PE). Introduction of PE post-assembly of these proteins results in complete reversal of the aberrant topological organization for PheP (Zhang et al., 2003) and at least the cytoplasmic domain C6 of LacY (Bogdanov et al., 2002). The above permeases maintain a compact folded state in the absence of PE as indicated by retention of energy independent downhill

transport function and resistance to degradation (Bogdanov and Dowhan, 1995; Bogdanov et al., 2002; Zhang et al., 2003, 2005).

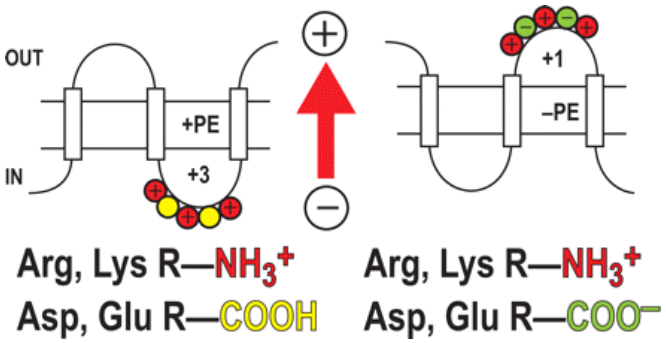


Figure 17. PE and orientation of membrane proteins in bacteria. In the left panel a cytoplasmic domain is shown containing a mixture of negative and positive amino acids. PE suppresses or neutralizes the presence of negative residues (yellow circles), which increases the effective positive charge potential, thus favoring cytoplasmic retention of the domain. In the absence of PE (right panel) negative residues (green circles) exert their full potential, resulting in translocation of the domain with a lower effective net positive charge. The proton motive force (arrow) positive outward determines domain directional movement depending on the domain effective net charge as influenced by the lipid environment. Taken from (Bogdanov et al., 2002)

i.3.2.6. Lateral pressure in the membrane

Lipid molecules have different geometries, depending on the size of the head group and how saturated their hydrocarbon tails are. Simply put, lipids shaped as cylinders are bilayer-forming and lipids shaped like cones introduce curvature (Marsh, 2006). If non-bilayer lipids are forced into the bilayer a stress is introduced in the membrane. The stress is also affected by the charge states of the lipid headgroups (Booth and Curran, 1999). The lateral pressure profile through the membrane, or membrane curvature, differs with lipid composition, and affects the folding and insertion of some proteins (**Figure 18**).

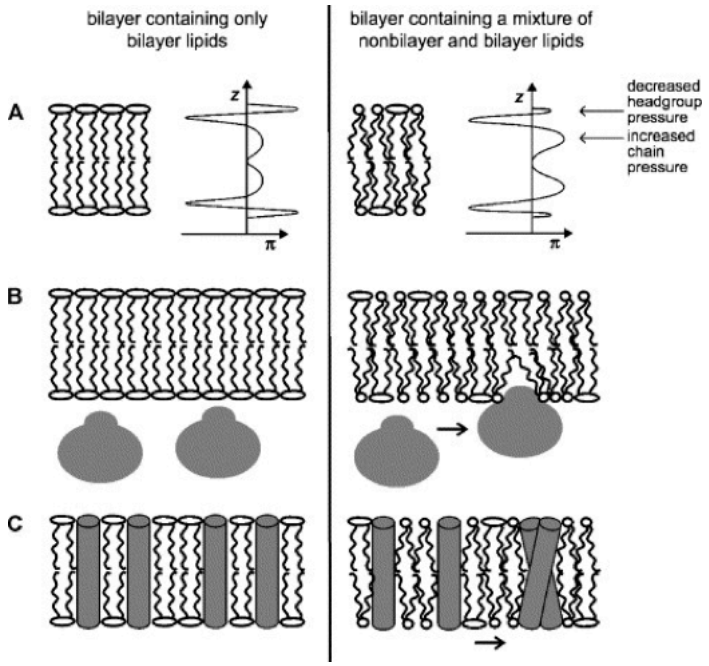


Figure 18. Predictions on the effect of nonbilayer lipids on peripheral and integral (multi-spanning) membrane proteins via the lateral pressure profile. (A) The lateral pressure profile is indicated for a bilayer containing either only bilayer lipids (left) or a mixture of nonbilayer and bilayer lipids (right). The effect on both peripheral membrane proteins (B) and multi-spanning integral membrane proteins (C) is indicated below. In extreme: nonbilayer lipids are expected to support membrane binding of peripheral membrane proteins and stabilize the oligomeric structure of integral membrane proteins. (Curran et al., 1999)

i.3.3. Hydrophobic mismatch between membrane and protein

Hydrophobic mismatch, which is the difference between the hydrophobic length of TM segments of a protein and the hydrophobic width of the surrounding lipid bilayer, is known to play a role in membrane protein function. It has been proposed that it is energetically favorable for the membrane protein to match the hydrophobic thickness of the lipid bilayer with a similar length of hydrophobic domain of the peptide (de Planque and Killian, 2003; Jensen and Mouritsen, 2004; Killian, 1998; Mouritsen and Bloom, 1993). When a hydrophobic mismatch exists, the peptide-lipid system undergoes compensatory adjustments to mitigate the energetically unfavorable mismatch in lengths. Such adjustments in structure or orientation of peptides or lipids could play important roles in protein activity. The lengths of the lipid hydrocarbon tails and the TM segments of membrane proteins are not always matched, **Figure 19**.

To alleviate hydrophobic mismatch, TM peptides and lipids can undergo adaptations, such as those shown in **Figure 19**. For positive hydrophobic mismatch, i.e., a protein hydrophobic length that is greater than the thickness of the lipid hydrophobic region, one or more of the following adaptations can occur:

1. The α-helix can reduce its hydrophobic length by becoming a less tightly bound π-helix-like conformation.
2. The peptide can tilt, reducing its exposure to polar groups.
3. The acyl chains near the peptide can order, increasing the local bilayer hydrophobic width by approaching phospholipids with longer acyl chains.
4. The peptides can oligomerize or aggregate.
5. The peptide can be expelled from the bilayer.

In case of a negative hydrophobic mismatch, one or more of the following adaptations can occur:

1. The α-helix can increase its hydrophobic length by becoming a more tightly bound 3_{10} helix-like conformation.
2. The bilayer width near the peptide can decrease, by acylchain disordering.
3. The peptides may aggregate or oligomerize.
4. The peptide can induce nonlamellar phase formation.
5. The peptide can be expelled from the bilayer.

Moreover, mismatch has been suggested to affect the functionality of membrane proteins, partly due to effects on TM helix packing (Orzáez et al., 2005). Eukaryotic cells have also been suggested to utilize hydrophobic mismatch to sort membrane proteins in the exocytic pathway (Sharpe et al., 2010).

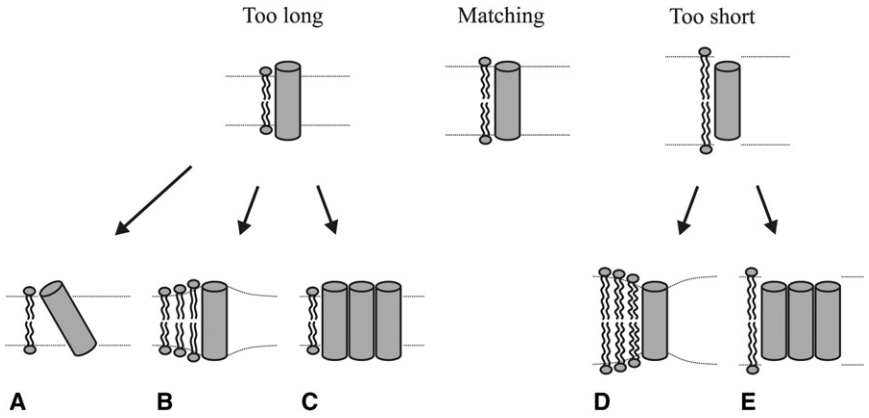


Figure 19. Possible adaptations to hydrophobic mismatch. In the case of too-long transmembrane peptides: (A) peptide tilting, (B) bilayer distortion, and/or (C) peptide aggregation). For too-short transmembrane peptides: (D) bilayer distortion and/or (E) peptide aggregation). Taken from (Ramadurai et al., 2010)

I.4. How is topology decided during biogenesis and folding?

Deciphering the topology signals and folding of a membrane protein are often described as separate processes in a simplified two-step model (Popot and Engelman, 1990; Popot et al., 1987). In this model, topology is established first, as the helices are thought to insert individually and to be stable on their own in the membrane. Thereafter the helices are envisioned to interact with each other within the membrane and form the final tertiary structure. Although this model may be true for some proteins (Popot et al., 1987), it is too simplified for others. There are several examples on helices of low hydrophobicity, marginally hydrophobic helices, which are not inserting well on their own, but are TM in the folded protein; (Hedin et al., 2010; Heinrich and Rapoport, 2003; Ota et al., 1998; Tamborero et al., 2011). During the folding of aquaporin1, the prospective TM helix2 and TM helix 4 are even extramembrane after passing the translocon and are pulled into the membrane as TMH3 later flips 180 degrees, assisted by the nascent C-terminal helix bundle (Pitonzo et al., 2009a). Other examples of double-spanning membrane proteins have been shown the two TM helices leaving the translocon and entering into the lipid bilayer together (Saurí et al., 2005; 2007). In addition, it is known that some TM helices interact transiently with proteins associated with the pore upon entering into the lipid bilayer. Emerging evidence also show that helices can stay in close proximity to the translocon until stable

interactions can be formed with subsequent helices (Sadlish et al., 2005) and that some of them can even be retained within the channel due to specific interactions (Pitonzo et al., 2009a).

Still, the three major factors affecting how a peptide chain is threaded through the membrane seem to be hydrophobicity, length of the hydrophobic segment and the presence of charged residues in the loops. For some segments, the signals to adopt a certain topology can be so strong that they force an intermediate segment to become TM, despite its low hydrophobicity (Ota et al., 1998). Timing of the elongation and insertion can also affect topology, as the proper segments need to be available at the correct moment, and so on (Goder et al., 2004).

i.4.1. Charged residues in loops, the "positive inside" rule

Positively charged, basic residues are strong topogenic signals with a preference for the cytosol. A simple way to predict the topology of a membrane protein is to localize the most hydrophobic parts (putative TM segments) and then summarize the number of positively charged residues in the loops on either side of the membrane. The side presenting the highest number of positively charged residues is likely to face the cytosol (Heijne, 1986). This effect is called the "positive inside rule" (Heijne, 1986) and has been shown to hold for most organisms (Heijne and Gavel, 1988). Acidic residues do not affect topology as strongly and show no statistical preference for loops on either side of the membrane (Andersson et al., 1992; Baeza-Delgado et al., 2013). However, they seem to have a greater influence on topology in Sec-independent translocation (Rutz et al., 1999).

i.4.1.1. Possible mechanisms underlying the positive inside rule

The exact mechanisms behind the positive inside rule are not yet fully understood, and they may differ slightly between archaea, bacteria and eukarya. For example, the retaining cytosolic effect of positive charges is more pronounced in *E.coli* than in mammalian microsomes (Johansson et al., 1993), most likely due to the much stronger electrochemical potential across the bacterial inner cell membrane. The ER membrane potential seems to be non-existing. In addition, in *E. coli*, the dominance of positive charges over negative charges as determinants of topology has been proposed to be due to the lipid composition

of the cell membrane. Several *E. coli* proteins have been shown to be dependent on the presence of lipids with a net zero charge to achieve the proper topology (Bogdanov et al., 2002). As mentioned above, is thought that these lipids dilute the otherwise strong negative charge of the lipid head groups so that the acidic residues (aspartic and glutamic acid) stay protonized and thus uncharged, see **Figure 17**. Specific interactions with the translocon have been also reported to contribute to the orientation of the first TM segment, the signal sequence (Goder et al., 2004; Junne et al., 2007). Finally, electrostatic interactions between anionic lipids and positively charged residues have been proposed to help retaining positive charges in the cytosol (van Dalen and de Kruijff, 2004; van Klompenburg et al., 1997).

i.4.2. Hydrophobicity as recognized by the translocon

Traditionally, the hydrophobicity of an amino acid has been determined by its ability to partition between water and an organic solvent (Jacobs and White, 1989). A subsequent improved method includes the energetically important effect of the protein backbone on insertion (White and Wimley, 1999; Wimley and White, 1996; Wimley et al., 1996). More recently, has been proposed a biological hydrophobicity scale (Hessa et al., 2005). In this scale, rather than looking at the hydrophobicity of the individual amino acids or side chains by partition experiments, the authors provide information on how each side chain will affect the recognition of a helix as TM by the translocon (**Figure 20**). $\Delta G_{app} \leq 0$ kcal/mol means that the amino acid favors recognition by the translocon. Positional effects are also taken into consideration (Hessa et al., 2007). For the work presented in this thesis, these biological scale and tools for predicting the change in free energy upon insertion based on this scale were employed.

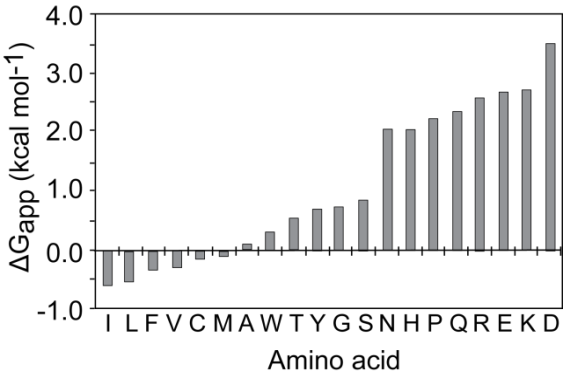


Figure 20. A biological scale. Apparent change in free energy (ΔG_{app}) upon translocon-mediated insertion into the ER membrane. Adapted from (Hessa et al., 2005).

i.4.3. Structure analysis

The function of a protein depends on its structure and folding. The three-dimensional structures of proteins are thus of uttermost interest. For example, once TM protein segments were thought to be perfect α-helices, long enough to span the lipid bilayer. As more 3D-structures of MPs have been obtained, it has become clear that TM elements of higher complexity are commonly occurring, including the previously described marginally hydrophobic TMHs. This knowledge allows us to better understand the mechanisms behind protein functionality and the effect of the primary sequence on folding. For the final structure of membrane proteins, topology is a vital part. There are three ways to attain information on topology: (1) Solve the three-dimensional structure. This is especially challenging for membrane proteins, but gives detailed information on all aspects of the structure, not only the topology. (2) Investigate the topology by biochemical means. (3) Predict the topology of a protein of interest based on previous knowledge. A number of *in silico* topology predictors have been developed as the collective dataset has grown during the last decades (Bernsel et al., 2008; Heijne, 1992; Käll et al., 2004; Krogh et al., 2001; Tusnády and Simon, 2001; Viklund et al., 2008). However, so far there are no tools integrating predictions of topology, coils, helix packing, surface accessibility, and so on, into a detailed prediction of membrane protein structure (Nugent and Jones, 2010). Many more experimental analyses of membrane protein structures are clearly needed to improve predictions.

i.4.4. Determining membrane protein topology experimentally

How can the topology of a membrane protein be determined? One way is to utilize a modification that can occur in only one of the compartments separated by the membrane, like glycosylation (Martínez-Gil et al., 2011), or hydrophilic cysteine labeling (Bogdanov et al., 2002). By gradually introducing acceptor sites for the modification tag at different positions and analyzing if the modification comes about or not, the topology can be deciphered. Reporter proteins that will be active only on one side of the membrane is a variation on the same theme like phosphatase A experiments in bacteria, but can in general only give information on the location of the protein termini. Another approach is "membrane shaving", where all extramembrane protein loops are cut into pieces by a peptidase (Wu and Yates, 2003). The individual peptides in the resulting mix are then identified using mass spectrometry (Bendz et al., 2013), which might require specific modification of the peptides (Jansson et al., 2008), depending on the protease used.

i.4.4.1. *N*-linked glycosylation as topology reporter and molecular ruler

In eukaryotic cells, the enzyme OligoSaccharyl Transferase (OST) is found on the lumenal side of the ER membrane (**Figure 21**). It recognizes the peptide sequence NX(S/T) and will attach one sugar moiety to the asparagine residue. The glycosylation process bassed on OST activity is co-translational. The attached oligosaccharide increases the molecular mass of the protein with approximately 2 kDa, and glycosylated and non-glycosylated forms of the protein can easily be separated by SDS-PAGE. In the recognition sequence, X cannot be Pro (Shakin-Eshleman et al., 1996). NXT is preferred over NXS, as the recognition of the latter sequon is more affected by the residue following the hydroxilated Thr/Ser residue (Kasturi et al., 1997; Mellquist et al., 1998; Shakin-Eshleman et al., 1996).

The active site of the transferase is located at a fixed distance from the membrane. Half-maximal glycosylation efficiency will be achieved if the acceptor site is (a) at least 14 residues N-terminal of a TM segment or (b) at least 12 residues C-terminal of the most C-terminal TM residue (Nilsson and Heijne, 1993b). The position of a TM segment can thus be precisely determined by finding the "minimal glycosylation distance" (MGD), the position where the acceptor asparagine is glycosylated at half-maximal efficiency (Monné et al., 1998; Nilsson and Heijne, 1993a; Stefansson et al., 2004).

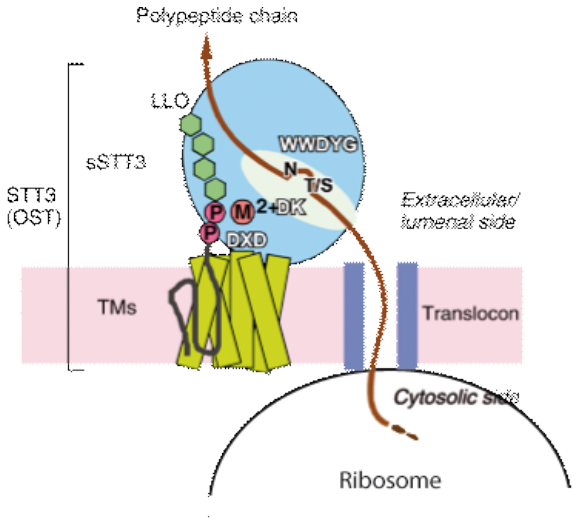


Figure 21. Schematic of the overall structure of OST. The overall structure of the full-length OST, and the location of the three conserved motifs, WWDYG, DK, and DXD. A possible binding mode of a ribosome nascent polypeptide chain emerged from a translocon, and that of an oligosaccharide-PP-dolichol/undecaprenol are shown. The aspartate in the WWDYG motif functions as a catalytic base. The DK and DXD motifs are the binding sites for the pyrophosphate group of lipid-linked oligosaccharide donor, through a transiently bound Mn2+or Mg+, in a coordinated or sequential manner. Adapted from (Igura et al., 2008).

i.4.5. Determining the propensity to integrate into the membrane

The main method used in this thesis to determine insertion propensity of a protein segment employs *N*-linked glycosylation as a reporter. A model protein carrying the segment of interest and two glycosylation acceptor sites is expressed in an *in vitro* system. Non-glycosylated, singly and doubly glycosylated protein species are subsequentially separated by SDS-PAGE and the signal intensities from the radioactive labeled bands on the gel are quantified. The insertion propensity is reported as the fraction of inserted protein species and is determined by the glycosylation status of the protein construct. The apparent change in Gibbs free energy upon insertion (ΔG_{app}) can be calculated for this system, using a simplified model where the translocon-mediated integration, the partioning of the protein

segment between the aqueous pore of the translocon and the hydrophobic bilayer, is regarded purely as an equilibrium process (Hessa et al., 2007).

$$K_{eq} = \frac{f_{inserted}}{f_{not\ inserted}} = \frac{f_{inserted}}{1 - f_{inserted}}$$

$$\Delta G_{app} = -RT \times \ln(K_{eq})$$

Where K_{eq} is the equilibrium constant, $f_{inserted}$ is the fraction of protein species where the segment was inserted into the membrane, R is the gas constant and T is the absolute temperature.

i.4.6. Model proteins, Lep and Lep´

As mentioned earlier, a highly hydrophobic signal peptide is required for a nascent polypeptide to enter the translocon-mediated insertion pathway. To be able to determine how well a less hydrophobic segment inserts into the membrane, either the entire native protein is used or the segment of interest can be isolated by fusing it to a model protein. In this work, the mainly used model proteins were two modified versions of the well-studied *E. coli* membrane protein Leader peptidase B (Uniprot P00803), named Lep and Lep´. Leader peptidase (Lep) is a protein of 324 amino acid residues that contains two TM segments (H1 and H2). The first 35 residues of leader peptidase, which include the first hydrophobic domain and the carboxyl-terminal positively charged cluster, are sufficient to insert the amino terminus with an N-terminal extracytoplasmic orientation (Sung and Dalbey, 1992). When positively charged residues are introduced before the first transmembrane segment, translocation of the amino-terminus is abolished (Andersson et al., 1992) . Both its short N-terminus and the large C-terminal domain face the lumen when expressed in an *in vitro* system of rabbit reticulocyte lysate (Johansson et al., 1993; Nilsson and Heijne, 1993a). In Lep, the fused segment to be tested is inserted within the large C-terminal domain (P2). In Lep´ it replaces the second TM segment (H2). The N-terminal of the segment can thus be positioned either toward the lumen (N_{out}, in Lep) or in the cytosol (N_{in}, in Lep´) depending on which model protein analysed, see **Figure 22**. Both model proteins carry two glycosylation sites, placed on appropriate distances from the lipid bilayer in a manner such that the location of the fused segment (inserted/non-inserted) is mirrored by the glycosylation status (see **Figure 22**A).

To study membrane protein regions containing several TM segments, extra glycosylation sites can be introduced by altering putative extra-membrane sequences of the foreign regions.

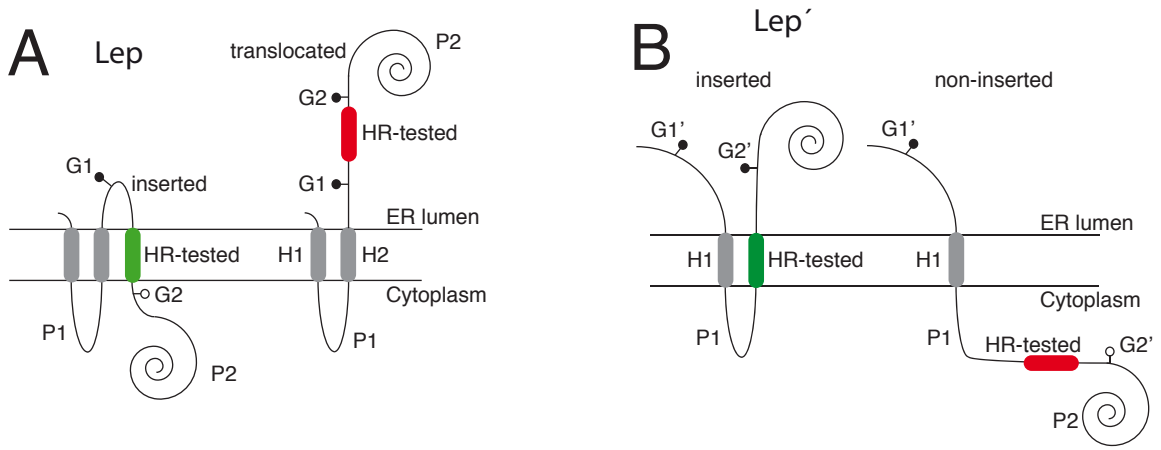


Figure 22. Lep model proteins. Grey rods (H1 and H2), native TM helices of LepB. Coloured rod, segment to be investigated. G1, G2, G1´ and G2´ positions for glycosylation recognition sites. The sites recieve a glycan (•) if translocated across the membrane to the lumenal side. A) Lep. One glycosylation indicates insertion of the segment of interest. (B) Lep´,, Two glycosylations indicate integration of the segment of interest into the membrane. Note that the N-terminal region at Lep´ is longer than for Lep construct, in order to ensure glycosylation of site G1´.

i.4.7. *In vitro* protein expression

Intact cells are not needed for protein expression. Cell lysates containing all the necessary components are enough, and facilitates labeling and purification of studied proteins. During the last four decades a number of *in vitro* systems have been made commercially available, prokaryotic as well as eukaryotic. They all contain a transcription/translation system as well as the signal peptide receptor system. When supplied with a polymerase of choice, some of these systems can conveniently transcribe and translate a protein directly from a DNA plasmid or even from a PCR fragment under the control of the appropriate promoter sequence. To facilitate a eukaryotic ribosome binding to the transcribed mRNA and ensuring a high rate of translation, it is important to include the Kozak consensus sequence (Kozak, 1989) upstream of the gene to be expressed.

In a lysate *in vitro* system, many proteins are already present. To facilitate visualization of the newly synthesized proteins of interest, expression can be carried out using radioactive amino acids, followed by proper visualization methods. Radioactivity also facilitates quantification of the expressed proteins (see above).

To study the biogenesis of membrane proteins, the system is supplemented with rough microsomes, small vesicular fragments derived from rough ER membranes, which contain all the translocon components need it for proper membrane integration.

PUBLICATIONS

Chapter 1:

N-Glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon

Article 1

N-Glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon

Manuel Bañó-Polo,[1] Francesca Baldin,[1] Silvia Tamborero,[1] Marc A. Marti-Renom,[2] and Ismael Mingarro[1]*

[1]Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot E-46100, València, Spain
[2]Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain

Received 13 July 2010; Revised 26 September 2010; Accepted 31 October 2010
DOI: 10.1002/pro.551
Published online 16 November 2010 proteinscience.org

Abstract: *N*-glycosylation is the most common and versatile protein modification. In eukaryotic cells, this modification is catalyzed cotranslationally by the enzyme oligosaccharyltransferase, which targets the β-amide of the asparagine in an Asn-Xaa-Ser/Thr consensus sequon (where Xaa is any amino acid but proline) in nascent proteins as they enter the endoplasmic reticulum. Because modification of the glycosylation acceptor site on membrane proteins occurs in a compartment-specific manner, the presence of glycosylation is used to indicate membrane protein topology. Moreover, glycosylation sites can be added to gain topological information. In this study, we explored the determinants of *N*-glycosylation with the *in vitro* transcription/translation of a truncated model protein in the presence of microsomes and surveyed 25,488 glycoproteins, of which 2,533 glycosylation sites had been experimentally validated. We found that glycosylation efficiency was dependent on both the distance to the C-terminus and the nature of the amino acid that preceded the consensus sequon. These findings establish a broadly applicable method for membrane protein tagging in topological studies.

Keywords: C-terminus tagging; glycosylation efficiency; membrane protein topology; oligosaccharyltransferase acceptor site; sequon

Introduction

Membrane proteins represent about a third of the proteins in all living organisms, but structural information is lacking for an understanding of their various functions. Based on the membrane proteins with 3D structures in the membrane protein database maintained in Stephen White's laboratory (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html), at the end of 2009, there were 217 membrane proteins with unique structures. These represented <0.4% of the total protein structures deposited in the Protein Data Bank.[1] Compared to soluble proteins, there is a striking paucity of membrane protein structures. Therefore, membrane protein topology (i.e., the number of transmembrane (TM)

Abbreviations: bR, bacteriorhodopsin; ER, endoplasmic reticulum; Lep, leader peptidase; RMs, rough microsomal membranes; SDS-PAGE, sodium dodecylsulfate polyacrylamide-gel electrophoresis; TM, transmembrane.

Grant sponsor: Spanish Ministry of Science and Innovation; Grant number: BFU2009-08401, BIO2007-66670, BFU2010-19310; Grant sponsor: Generalitat Valenciana; Grant numbers: ACOMP/10/247, PROMETEO/2010/005; Grant sponsors: MICINN, ERDF (EU); V Segles program (University of Valencia); Spanish Ministry of Education (FPU fellowship).

**Correspondence to:* I. Mingarro, Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot E-46100, Spain. E-mail: Ismael.Mingarro@uv.es

segments and their orientation in the membrane) provides an important intermediate picture, that is, more informative than the amino acid sequence,[2] although less than the fully folded 3D structure.

Our knowledge of the underpinnings of membrane protein structure has grown exponentially in the last few years.[3,4] Common membrane protein architectural features are necessary for insertion into the lipid environment of the cell membrane. Hence, the great majority of membrane proteins contain one or more TM α-helices formed by a stretch of ~20 amino acids with hydrophobic side chains. These hydrophobic TM regions are connected with hydrophilic loops with distinct charge distributions.[5] This provides a simple method for predicting the topology of a membrane protein, which is typically confirmed with molecular and biochemical techniques.

In eukaryotic cells, most membrane proteins are integrated into the membrane cotranslationally; that is, at the same time that they are being synthesized by ribosomes. They are incorporated into the endoplasmic reticulum (ER) membrane at sites termed translocons, which comprise a specific set of membrane proteins.[6] During this process, the translocon mediates the integration of TM sequences into the nonpolar core of the membrane bilayer and delivers hydrophilic cytoplasmic and luminal domains to the appropriate compartments. Simultaneously, a nascent protein may undergo covalent modifications, like signal sequence cleavage and *N*-glycosylation. *N*-glycosylation is performed in the lumen of the ER by the enzyme oligosaccharyltransferase (OST). OST transfers preassembled sugar moieties from a lipid carrier to the β-amino groups of the asparagine residues in the Asn-Xaa-Ser/Thr (NXS/T) consensus sequences.[7] Modifications of the glycosylation acceptor sites occur in a compartment-specific manner; thus, the presence of glycosylation can provide valuable topological information.[8] This endogenous glycosylation information can be extended experimentally by adding glycosylation tags at the C-terminus of the polypeptide. The aim of this study was to explore the determinants of glycosylation efficiency for added C-terminal tags. Our results showed that a C-terminal tag requires at least six amino acid residues for efficient glycosylation, and that the amino acid preceding the NXS/T sequon is an important determinant of glycosylation efficiency.

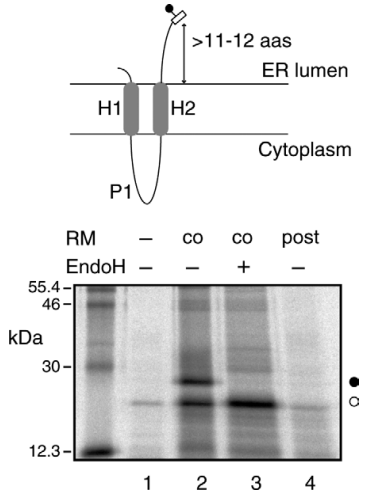


Figure 1. C-terminal-tagged truncated Lep constructs are cotranslationally glycosylated. (Top) Diagram showing the orientation of truncated Lep in the microsomal membrane. (Bottom) C-terminal-tagged (**NST**MMS) Lep construct was translated in either the absence (lanes 1 and 4) or presence (lanes 2 and 3) of dog pancreas rough microsomes (RM). After translation, samples were treated with Endo H (lane 3). In lane 4, RMs were added post-translationally; Lep constructs underwent 1 h translation, followed by 10 min cycloheximide treatment, then incubation with RMs was continued for another 1 h. Bands of unglycosylated and glycosylated proteins are indicated with white and black dots, respectively.

Results and Discussion

Glycosylation efficiency increases with distance from the C-terminus

To examine the influence of the distance between the glycosylation site and the C-terminus of the polypeptide, we expressed a truncated protein in an *in vitro* translation/glycosylation system with or without added dog pancreas microsomes. We utilized the well-characterized *Escherichia coli* inner membrane protein leader peptidase (Lep) harboring an Asn-Ser-Thr sequon, which is a well-known glycosylation motif. Lep is anchored in the cytoplasmic membrane by two TM segments (H1 and H2) that are connected by a highly positively-charged cytoplasmic domain (P1), which drives membrane topology.[9] When Lep was translated/glycosylated *in vitro* in the presence of dog pancreas microsomes, it inserted into the microsomal membrane with both the N and C termini on the luminal side.[10,11] Previous studies have shown that, when an engineered *N*-glycosylation site was placed downstream of H2 at 11-12 residues distal to the hydrophobic end (Fig. 1, top), it was glycosylated upon correct insertion into the microsomal membrane.[10,12] Glycosylation of the molecule resulted in a 2.5 kDa increase in molecular mass relative to that of Lep expressed in the absence of microsomes. To determine whether glycosylation efficiency was affected by the position of the glycosylation acceptor site, we generated truncated proteins that included *N*-linked glycosylation sites at different distances from the C-terminus. These truncated Lep variants were expressed in a rabbit reticulocyte cell-free translation system supplemented with [^{35}S]

Met/Cys and dog pancreas rough microsomes. Translation of each variant yielded two types of protein products: the truncated Lep protein with a single oligosaccharide attached to the tag and the unglycosylated truncated protein. The proportion of glycosylated and unglycosylated proteins directly reflected the efficiency of *N*-glycosylation by OST. After SDS-PAGE analysis, the proportions of glycosylated and unglycosylated protein were quantified from gel autoradiographs.

We first determined whether truncated Lep proteins that carried a C-terminal glycosylation tag were cotranslationally glycosylated. It has long been reported that the tripeptide sequon Asn-Xaa-Thr is more efficiently glycosylated than Asn-Xaa-Ser;[13] in fact, the occurrence rate of the former is about one-third higher than that of the latter (39,161 and 30,579 sequons in our database, respectively), which is in agreement with a recent statistical survey.[14] We found that translation of truncated Lep with a six residue glycosylation tag (**NST**MMS) in the presence of rough microsomal membranes (RM) was associated with an increase in the molecular mass, indicative of protein glycosylation (Fig. 1, lane 2). This result was further corroborated when the increase in mass was abolished by treatment with endoglycosidase H (Fig. 1, lane 3), a glycan-removing enzyme.[15] Notably, when microsomal membranes were included post-translationally, after translation inhibition with cycloheximide, the C-terminal acceptor site was not glycosylated (Fig. 1, lane 4); this suggested that the truncated Lep was integrated into the membrane cotranslationally, via the ER translocon. These results were consistent with an earlier study on truncated Lep, where the glycosylation efficiency was reduced to ~40% when the glycosylation acceptor site was placed six residues upstream of the stop codon.[16] Similarly, it has been shown that the recombinant mammalian concentrative Na^+-nucleoside cotransporter rCNT1 could be glycosylated in *Xenopus* oocytes at an asparagine residue located six residues upstream of the C-terminal end.[17]

Next, we investigated glycosylation efficiency as a function of the distance between the acceptor Asn residue and the C-terminus of the polypeptide. As shown in Figure 2(A,B), the glycosylation efficiency increased gradually with the distance between the acceptor site and the C-terminus. When the C-terminal glycosylation tag only included the three amino acid residues that formed the acceptor sequon (**NST**, 3 residues tag), the truncated polypeptide remained unglycosylated [Fig. 2(A), lanes 1 and 2]. Extending the C-terminal tag to four residues (**NST**M) slightly increased glycosylation (~20%, lanes 3 and 4), and a C-terminal tag with five residues (**NST**MM) nearly doubled the glycosylation efficiency [Fig. 2(A), lanes 5 and 6]. Further extensions of the tag length

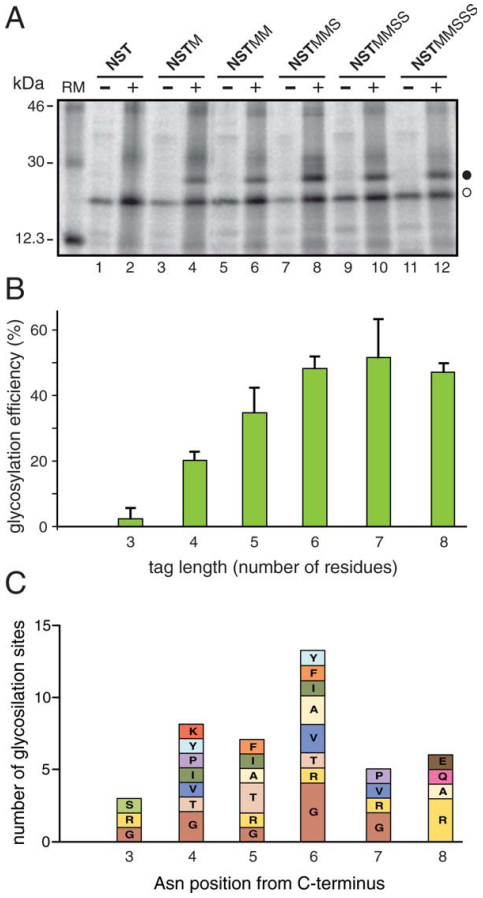


Figure 2. Glycosylation efficiency increases with the distance from the C-terminus. (A) *In vitro* translation of the truncated proteins with different C-terminal glycosylation tags in the absence (–) and in the presence (+) of RM. As in Figure 1, glycosylated and unglycosylated products are shown with black and white dots, respectively. (B) The glycosylation efficiency is shown as a function of the number of residues between the acceptor Asn and the C-terminus (tag length). To calculate the percent efficiencies, the total glycosylation (100%) was taken as the sum of the signals present in the glycosylated and nonglycosylated forms. Data correspond to averages of at least three independent experiments; error bars show standard deviations. (C) NXT glycosylation sequon distribution at the C-terminal region (positions 3 to 8) in nonredundant experimentally validated glycoproteins. Each bar height is proportional to the number of sequons and displays the distribution of amino acid residues at each position. Nonoccurring amino acid residues at each site are omitted.

rendered similar glycosylation levels [~50%, see Fig. 2(A), lanes 7-12]. To compare these results with native glycoproteins, we performed a statistical analysis using the sequences and their annotations from the UniProt database (http://www.uniprot.org,

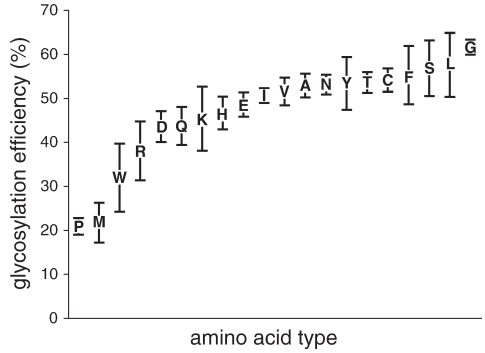


Figure 3. Glycosylation efficiencies of Lep truncates with different amino acid residues preceding the glycosylation sequon. C-terminal-tagged truncated Lep variants contained the indicated amino acid residues in front of the Asn residue of the glycosylation site. Glycosylation levels were determined from gel autoradiographs. Data correspond to averages of at least three independent experiments; error bars show standard deviations.

release 2010_09).[18] After selecting nonredundant *N*-glycosylated proteins (see Materials and Methods section), the complete set of putative *N*-glycosylation sites was obtained by selecting only Asn-Xaa-Thr sequons. The final dataset contained 39,161 sequons of which 5,753 were experimentally validated. Native glycosylated sites located at the C-terminal regions were more prominent (13 occurrences) for sequons with the Asn amino acid located six residues upstream from the C-terminus [Fig. 2(C)]. Nevertheless, the total number of glycosylation sites at this position relative to the total sequons (5,753 in native sequences) suggests that protein glycosylation near the C-terminus is a relatively rare event and explains the low glycosylation efficiency (~50%) in our experiments. Thus, the tag with six amino acid residues (**NST**MMS) was selected for further experiments. It should be noted that the presence of a methionine residue following the glycosylation sequon conferred optimal glycosylation efficiency when Thr was present at the hydroxyl (third) position.[19]

Glycosylation of truncated Lep variants

We translated 20 variants of C-terminal-tagged truncated Lep proteins to examine systematically whether the amino acid residue preceding the acceptor Asn affected glycosylation efficiency (Fig. 3). As expected, when a Pro residue preceded the acceptor Asn, glycosylation was significantly inhibited. However, Pro had a stronger inhibitory effect when it was located either at the central Xaa position[20] or following the glycosylation sequon.[19] It is interesting to note that Pro has never found preceding an experimentally verified glycosylation site in our database when the glycosylated Asn residue in the NXT sequon is located at six residues from the C-terminal end [Fig. 2(C)]. However, this inhibitory effect was not observed when the Pro residue was inserted just before the acceptor Asn in a full-length Lep construct (Fig. 4). In fact, more than 80% of the molecules were glycosylated when this Lep mutant was assayed (Fig. 4, lane 8). This suggested that the residue preceding the glycosylation sequons only impacted glycosylation efficiency when the acceptor Asn residue was close to the end of the polypeptide. Indeed, of the 42 sequons Asn-Xaa-Thr located within the last eight residues, only two were preceded a Pro residue [Fig. 2(C)].

The probability of each amino acid type preceding a verified glycosylation sites has been calculated for the Asn-Xaa-Thr sequons in the nonredundant dataset (Fig. 5). All 20 amino acids can be found preceding the Asn residue of the sequons, although significant differences between their probabilities occur in the experimentally validated glycosylation sites. Experimentally, we found that the glycosylation efficiency of the NST sequon was also significantly lowered when it was preceded by Met, Trp, or Arg residues (Fig. 3), which correlates with the results of our statistical analysis, especially in the case of Met and Trp (Fig. 5). One explanation for this observation

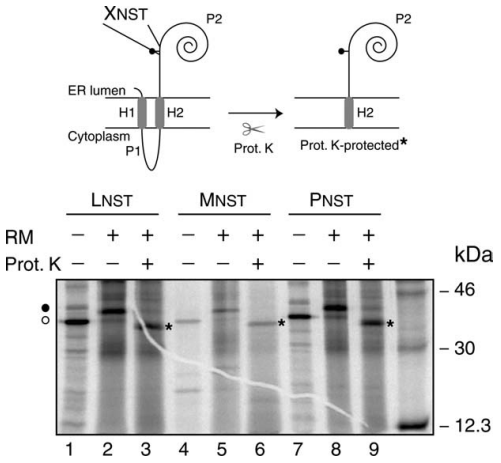


Figure 4. Glycosylation efficiency of full-length Lep mutants. *In vitro* translation of mRNAs encoding full-length Lep mutants was achieved in the presence (+) and absence (–) of membranes and proteinase K (PK) as indicated. Lep variants contain a single Asn-Ser-Thr sequon (codons 97-99) preceded by Leu (lanes 1-3), Met (lanes 4-6) or Pro (lanes 7-9) in each case. Bands of nonglycosylated protein are indicated by a white dot and glycosylated proteins are indicated by a black dot. The asterisk identifies undigested protein after PK treatment. (Top) Schematic representations of the Lep full-length construct and the proteinase K-protected fragment.

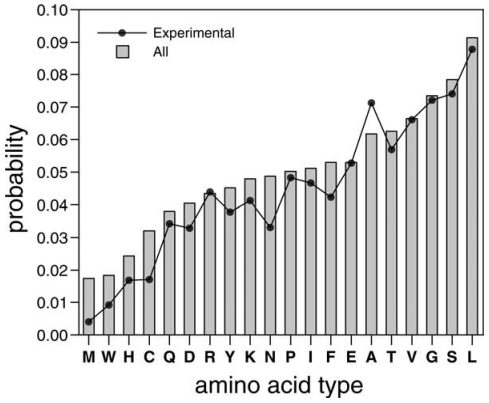


Figure 5. Distribution of amino acid residues preceding NXT glycosylation sequons in all sites (gray bars) and experimentally validated sites (black line).

might be that the bulky side chains of these residues may block accessibility to the OST active site or the lipid carrier donor; another explanation could be that it may induce an unfavorable local protein conformation. Previous studies revealed that glycosylation was strongly inhibited when Trp was placed at the central Xaa position,[20,21] and it was somewhat inhibited when Trp followed the glycosylation sequon.[19] Our results also pointed out some average effect caused by the presence of acidic residues immediately before the glycosylation site. It is interesting to note that it has been described a notable reduction in the probability of finding acidic residues preceding the glycosylation site.[22] Even more, this is accompanied by an increase probability of finding acidic residues preceding unoccupied glycosylation sites.[14] However, both Asp and Glu have been found as average preceding NXT acceptor sites (Fig. 5). The apparent discrepancies between our and the previous studies could arise from the fact that the later surveys included the glycosylation sites with Ser in the third position (NXS/T), whilst our database focussed in NXT glycosylation sites. The other amino acid residues appeared to have only minor effects on glycosylation efficiency; however, the Gly residue consistently induced higher glycosylation levels (Fig. 3), and again, an increased probability of finding Gly preceding glycosylation sites was observed in our analysis (Fig. 5). We assumed that the flexibility that Gly confers on the conformation of the polypeptide chain may provide an advantage for OST catalysis. In fact, the structural conformation of the local region around the glycosylation sequon also influenced its accessibility and, consequently, its site occupancy.[23] This supports the hypothesis that unfolding or flexibility is required for protein domains to be efficiently glycosylated. It should be also noted that there is a marked preference for hydrophobic amino acids immediately preceding the glycosylation site in our experimental data, which nicely correlates with previous[22] and the present statistical analysis of glycan-protein linkage, especially in the case of Leu that has been also found prevalent in glycosylated sequons (see reference 14 and Fig. 5).

Interestingly, recent statistical analyses of active bacterial *N*-glycosylation site consensus sequences showed that Asn, Phe, Ser, and Leu residues frequently precede the acceptor Asn.[24] In the present study, we found that these same residues and Gly were the best suited for glycosylation in a eukaryote OST. Based on these results, we propose that bacterial and eukaryotic systems might require similar sequences flanking the acceptor sites to adopt an optimal conformation upon the binding of OST.

Statistical studies have also shown that the glycosylation sequon occurs at the C-terminal end of well-defined glycoproteins at a lower frequency than that expected based on random chance;[13,14] and that when *N*-glycosylation sites are contained within more than one extracytosolic loop, only the first loop is modified.[25] Furthermore, those studies found that the glycosylation efficiency for Asn-Xaa-Thr sequons dropped when located close to the C-terminal end of the protein.[14] The present work pointed out that this effect was emphasized at the very end of the protein [Fig. 2(C)]. This suggested that it was necessary to use sufficiently large C-terminal glycosylation tags. In fact, we found that at least six residues long C-terminal glycosylation tags were needed to achieve significant glycosylation; this validated their utility in membrane protein topological studies.

To prove our approach, we have fused the N-terminus of bacteriorhodopsin (bR) (from Trp10 to Val101) at the C-terminus of the engineered Lep sequence (see Materials and Methods section). We choose bR because it is a membrane protein with a well-defined topology, in which the N-terminus faces the extracellular side similarly to our chimeric constructs (Fig. 6, top). *In vitro* transcription/translation of protein truncates using a glycosylable C-terminal tag after bR helix a (the first TM segment) rendered singly-glycosylated forms (Fig. 6, lane 2), indicating the insertion of bR helix a. Truncated polypeptides, which include the first two TM helices of bR, were efficiently doubly-glycosylated (Fig. 6, lane 4), demonstrating translocation of the glycosylation site included as a C-terminal tag, and validating our experimental approach.

Conclusion

We have investigated N-linked glycosylation efficiency using an *in vitro* system based on microsomes and a well-characterized model protein. In conclusion, we found that in placing a glycosylation tag on a

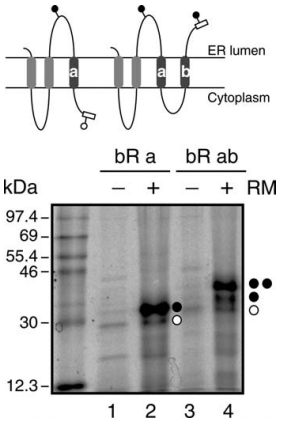


Figure 6. Glycosylation efficiency of Lep/bR truncates. *In vitro* translation of C-terminal-tagged mRNAs encoding Lep/bR constructs was performed in the presence (+) and absence (–) of membranes as indicated. Bands of nonglycosylated proteins are indicated by a white dot and singly- and doubly-glycosylated proteins are indicated by one and two black dots, respectively. (Top) Schematic representations of the Lep/bR constructs including bR helix a (left) and bR helices a and b (right).

polypeptide chain, one should consider both the distance from the hydrophobic end of a TM segment and the nature of the amino acid residue preceding the acceptor Asn residue. Taken together, our results provided a rapid and efficient method for the determination of membrane protein topology.

Materials and Methods

Enzymes and chemicals

The pGEM1 plasmid, rabbit reticulocyte lysate, and the TnT coupled transcription/translation system were purchased from Promega (Madison, WI). The ER rough microsomes from dog pancreas and the SP6 RNA polymerase were purchased from tRNA Probes (College Station, TX). The [^{35}S] Met/Cys and ^{14}C-methylated markers were purchased from Perkin-Elmer. The restriction enzymes and endoglycosidase H were purchased from Roche Molecular Biochemicals. The DNA plasmid, RNA clean up, and PCR purification kits were from Qiagen (Hilden, Germany). The PCR mutagenesis kit, QuikChange was from Stratagene (La Jolla, CA). All the oligonucleotides were purchased from Thermo (Ulm, Germany).

DNA manipulations

Full-length Lep DNA was amplified directly from the pGEM1 plasmid, which carried a modified *lep* gene. In that sequence, the nucleotides that encoded the Asn-Glu-Thr glycosylation acceptor site at position 214-216 in the wild type protein was changed to a nonacceptor sequence Gln-Glu-Thr. In addition, an Asn-Ser-Thr (NST) glycosylation acceptor site was introduced 20 amino acids downstream of H2 at codons 97-99. Alternatively, we prepared templates for *in vitro* transcription of the truncated wild type *lep* mRNA with a 3′ glycosylation tag. The truncated *lep* sequence was prepared by PCR amplification of a fragment of the pGEM1 plasmid that encoded the N-terminal 178 amino acid residues of Lep. The 5′ primer was the same for all PCR reactions and had the sequence 5′-TTCGTCCAACCAAACCGACTC-3′. This primer was situated 210 bases upstream of the *lep* translational start codon; thus, all amplified fragments contained the SP6 transcriptional promoter from pGEM1. The 3′ primers were designed to have approximately the same annealing temperature as the 5′ primer. They contained a glycosylation tag preceded by one of the 20 natural amino acids, and followed by the tandem translational stop codons, TAG and TAA. PCR amplification comprised a total of 30 cycles with an annealing temperature of 52°C. The amplified DNA products were purified with the Qiagen PCR purification kit, according to the manufacturer's protocol, and verified on a 1% agarose gel. The mutations Leu96 Met and Leu96Pro were performed with the QuikChange mutagenesis kit from Stratagene (La Jolla, CA), according to the manufacturer's protocol.

The N-terminal region from bacteriorhodopsin (residues 10-101) was PCR amplified and cloned into the modified Lep sequence from pGEM plasmid[26,27] between *Spe*I and *Kpn*I sites. The truncated Lep/bR chimeras were prepared by PCR amplification of fragments that encoded up to Lys41 (bR sequence) in the case of Lep/bRa and up to Ile78 for Lep/bRab truncates. All DNA manipulations were confirmed by sequencing the plasmid DNAs.

Expression in vitro

Truncated *lep* mRNAs with stop codons were transcribed from the SP6 promoter with SP6 RNA polymerase (tRNA probes). Briefly, the transcription mixture was incubated at 37°C for 2 h. The mRNA products were purified with a Qiagen RNeasy clean up kit and verified on a 1% agarose gel.

In vitro translation of *in vitro* transcribed mRNA was performed in the presence of reticulocyte lysate, [^{35}S] Met/Cys, and dog pancreas microsomes, as described previously.[28] For the posttranslational membrane insertion experiments, Lep-derived mRNAs were translated (30°C 1 h) in the absence of RMs. Translation was then inhibited with cycloheximide (10 min, 26°C, 2 mg/mL final concentration), after which RMs were added and incubated for an additional hour at 30°C.[29] In all cases, after translation, membranes were collected by ultra-centrifugation and analyzed by sodium-dodecylsulfate-polyacrylamide gel electrophoresis

(SDS-PAGE). Finally, the gels were visualized on a Fuji FLA3000 phosphorimager with ImageGauge software.

For endoglycosidase H (Endo H) treatment, the translation mixture was diluted in four volumes of 70 m*M* sodium citrate (pH 5.6) and centrifuged ($10^5 g$ for 20 min at 4°C). The pellet was then resuspended in 50 μL of sodium citrate buffer with 0.5% SDS and 1% β-mercaptoethanol, boiled for 5 min, and incubated for 1 h at 37°C with 0.1 mU of Endo H. The samples were analyzed by SDS-PAGE.

Full-length Lep constructs were transcribed and translated in the TnT Quick system (Promega). 1 μg DNA template, 1 μL 35S-Met/Cys (5 μCi) and 1 μL microsomes (tRNA Probes) were added at the start of the reaction, and samples were incubated for 90 min at 30°C. Translation products were analyzed as previously described for the truncated molecules. For the proteinase K protection assay, the translation mixture was supplemented with 1 μL of 50 m*M* $CaCl_2$ and 1 μL of proteinase K (4 mg/mL), then, digested for 40 min on ice. The reaction was stopped by adding 1 m*M* phenylmethanesulfonylfluoride before SDS-PAGE analysis.

Statistical analysis of N*-glycosilation sites in native proteins*

Sequences and their annotations were obtained from the UniProt database (http://www.uniprot.org, release 2010_09).[18] Selection of *N*-glycosylated proteins was done using the UniProt search engine by selecting all sequence annotation (*FT* field) as glycosylated modified amino acid. Such selection contained both, experimentally validated as well as nonvalidated glycosylation sites. The total number of sequences containing at least one glycosylation site was 25,488, of which 2,533 had been experimentally validated. Next, all the selected sequences were compared to each other using the *cd-hit* program with default parameters.[30] Redundant sequences at the 90% sequence identity cut-off were removed. Finally, only *N*-glycosylation sites with sequons NXT (being X any of the 20 amino acid types) were maintained. Our final dataset, which contained 39,161 NXT sequons of which 5,753 were experimentally validated, could be considered as an up-to-date set of nonredundant sequences with annotated NXT *N*-glycosylation sites.

Acknowledgments

The authors thank Prof. Esteve Padrós (Universitat Autònoma de Barcelona) for the sequence of bacteriorhodopsin.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.
2. von Heijne G (2006) Membrane-protein topology. Nat Rev Mol Cell Biol 7:909–918.
3. White SH (2004) The progress of membrane protein structure determination. Protein Sci 13:1948–1949.
4. White SH (2009) Biophysical dissection of membrane proteins. Nature 459:344–346.
5. von Heijne G (1992) Membrane protein structure prediction—hydrophobicity analysis and the positive-Inside rule. J Mol Biol 225:487–494.
6. Alder NN, Johnson AE (2004) Cotranslational membrane protein biogenesis at the endoplasmic reticulum. J Biol Chem 279:22787–22790.
7. Welply JK, Shenbagamurthi P, Lennarz WJ, Naider F (1983) Substrate recognition by oligosaccharyltransferases. Studies on glycosylation of modified asn-x-ser/thr tripeptides. J Biol Chem 258:11856–11863.
8. Cheung JC, Reithmeier RA (2007) Scanning N-glycosylation mutagenesis of membrane proteins. Methods 41: 451–459.
9. von Heijne G (1989) Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 341:456–458.
10. Nilsson I, von Heijne G (1993) Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J Biol Chem 268:5798–5801.
11. Martinez-Gil L, Sanchez-Navarro JA, Cruz A, Pallas V, Perez-Gil J, Mingarro I (2009) Plant virus cell-to-cell movement is not dependent on the transmembrane disposition of its movement protein. J Virol 83:5535–5543.
12. Orzaez M, Salgado J, Gimenez-Giner A, Perez-Paya E, Mingarro I (2004) Influence of proline residues in transmembrane helix packing. J Mol Biol 335:631–640.
13. Gavel Y, von Heijne G (1990) Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites—implications for protein engineering. Protein Eng 3:433–442.
14. Ben-Dor S, Esterman N, Rubin E, Sharon N (2004) Biases and complex patterns in the residues flanking protein *N*-glycosylation sites. Glycobiology 14:95–101.
15. Vilar M, Sauri A, Monne M, Marcos JF, von Heijne G, Perez-Paya E, Mingarro I (2002) Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. J Biol Chem 277:23447–23452.
16. Nilsson I, von Heijne G (2000) Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. J Biol Chem 275: 17338–17343.
17. Hamilton SR, Yao SY, Ingram JC, Hadden DA, Ritzel MW, Gallagher MP, Henderson PJ, Cass CE, Young JD, Baldwin SA (2001) Subcellular distribution and membrane topology of the mammalian concentrative Na+-nucleoside cotransporter rCNT1. J Biol Chem 276:27981–27988.
18. The UniProt Consortium, T.U. (2010). The universal protein resource (UniProt) in 2010. Nucleic Acids Res 38:D142–148.
19. Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH (1998) The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. Biochemistry 37: 6833–6837.
20. Shakin-Eshleman SH, Spitalnik SL, Kasturi L (1996) The amino acid at the X position of an Asn-X-ser sequon is an important determinant of N-linked core-glycosylation efficiency. J Biol Chem 271:6363–6366.
21. Kasturi L, Chen H, Shakin-Eshleman SH (1997) Regulation of N-linked core glycosylation: use of a site-

directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem J 323:415–419.
22. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology 14:103–114.
23. Jones J, Krag SS, Betenbaugh MJ (2005) Controlling N-linked glycan site occupancy. Biochim Biophys Acta 1726:121–137.
24. Kowarik M, Young NM, Numao S, Schulz BL, Hug I, Callewaert N, Mills DC, Watson DC, Hernandez M, Kelly JF, Wacker M, Aebi M. (2006) Definition of the bacterial N-glycosylation site consensus sequence. EMBO J 25: 1957–1966.
25. Landolt-Marticorena C, Reithmeier RAF (1994) Asparagine-linked oligosaccharides are localized to single extracytosolic segments in multi-span membrane glycoproteins. Biochem J 302:253–260.
26. 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.
27. Martinez-Gil L, Sauri A, Vilar M, Pallas V, Mingarro I (2007) Membrane insertion and topology of the p7B movement protein of Melon Necrotic Spot Virus (MNSV). Virology 367:348–357.
28. 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.
29. 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.
30. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.

Reviewing Process

Response to Reviewers

Reviewer #2:

1. "As reported in the original review, the N-glycosylation of C-terminal residues in the Lep protein has been examined by the von Heijne group who found that acceptor sites as close as 6 residue from the end can be N-glycosylated using a cell-free translation system. The current paper confirms this finding using the identical experimental system with the additional assessment of sites as close as three residues. This extends somewhat the previous findings, but does not provide significant new information."
As stated by the reviewer the current paper confirms previous results regarding N-glycosylation of C-terminal residues in the Lep protein, but it extends these preliminary results in a systematic manner by:

- mapping the glycosylation efficiency using tags differing by one amino acid at a time,
- investigating the influence of the 20 amino acid residues preceding the target Asn,
- demonstrating the utility of the C-terminal *N*-glycosylation method with a designed membrane protein, and
- performing an up-to-date statistical analysis of *N*-glycosylation sequons in native proteins

2. "The relevance of the current findings to native membrane proteins still needs to be established. Provide examples of native membrane proteins (e.g. CNT) that contain C-terminal N-glycosylation sites (how close to the C-terminus?) as these are rare. Reference 14 provides examples of utilized N-glycosylation sites within 10 residues of the C-terminus in secreted proteins, some as close as 5 residues! So it is already known that N-glycosylation can occur post-translationally very close to the C-terminus."
As suggested by the reviewer the occurrence of N-glycosylation sites are very rare. To investigate this fact we have performed a large-scale statistical analysis of N-glycosylation sites in native glycoproteins. The UniProt database contained at least one experimentally validated glycosylation site in 2,533 protein sequences. Next, we selected only N-glycosylation sites with sequons NXT from non-redundant sequences at 90% sequence identity. In our final dataset, with 5,753 experimentally validated NXT sequons, only 42 sequons (new Fig. 2C) were observed within the 8 C-terminal residues (<0,8%). Nevertheless, we demonstrated that C-terminal tags (carrying a glycosylation site) as short as six residues can be used as an efficient experimental approach to map membrane protein topologies.

3. "Scanning N-glycosylation mutagenesis has been applied routinely to study the topology of membrane proteins, so the authors are not proposing a novel method. It would be of interest to apply this method to truncated membrane proteins with acceptors sites engineered into hydrophilic positions at least 6 residues from the new C-termini. A demonstration of the utility to the C-terminal N-glysosylation method to a membrane protein would greatly enhance the appeal of this paper."
To demonstrate the utility of the C-terminal N-glycosylation method we have fused bacteriorhodopsin to the C-terminus of Lep (new Fig. 6). We have chosen bacteriorhodopsin because it is a model membrane protein with a well-known topology in which the first TM segment (helix *a*) is oriented with the N-terminus facing the extracellular side, as occurs in our Lep-derived constructs. In these set of experiments we found that C-terminal tags located at the hydrophilic loop connecting helices *a* and *b* of bacteriorhodopsin were not glycosylated, whilst when the tag was positioned after helix *b* the protein was efficiently glycosylated, validating our experimental approach.

4. "Again, I would make Fig. 2 a continuous curve and Fig. 3 (color not necessary) into a bar graph."
We have now included a new graph in Fig. 2 (panel C), which represents the occurrence of amino acid type preceding the NXT sequons in the C-terminus of native proteins. Such a representation requires a bar graph to show the distribution of amino acids at each position. Therefore, we kindly request to maintain Fig. 2B as a bar graph, which makes it easy to relate to the other two panels.
Regarding Fig. 3, we agree with the reviewer that color is not necessary. However, we believe that our representation conveys its information adequately as it is.

5. The addition of the results from two previous publications concerning the effect of the nature of the preceding residue (-1) on the utilization of N-glycosylation sites is welcome. Reference 20 indicates that aromatic residues are preferred at position –1 while small residues are not. This is in

contrast to the results reported in this paper where glycine is the best residue at this position and tryptophan is relatively poor. Reference 14 indicates that leucine at –1 is prevalent in glycosylated sequons in agreement with this paper, while glutamate at –1 is prevalent in non-glycosylated sequons and is an average acceptor in this study. So the relevance of the current findings using an artificial construct in a cell-free translation system to native membrane glycoproteins is not clearly established.

The prevalence of the more significant amino acid residues has been thoroughly discussed on the light of our statistical analysis (Fig. 5), which slightly differs from those in references 14 and 20. The main difference between those and our analysis is the exclusion of NXS sequons from our database. Briefly, we found that small residues like Gly and Ser are prevalent when only NXT sequons are taken into account, while large residues like Trp are clearly infrequent. On the contrary, Glu is the eighth more frequent residue preceding NXT sequons.

Chapter 2:

Membrane Integration of Poliovirus 2B Viroporin

Article 2

JOURNAL OF VIROLOGY, Nov. 2011, p. 11315–11324 Vol. 85, No. 21
0022-538X/11/$12.00 doi:10.1128/JVI.05421-11

Membrane Integration of Poliovirus 2B Viroporin▽

Luis Martínez-Gil,[1]† Manuel Bañó-Polo,[1]† Natalia Redondo,[2] Silvia Sánchez-Martínez,[3] José Luis Nieva,[3] Luis Carrasco,[2] and Ismael Mingarro[1]*

Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46100 Burjassot, Spain[1]; Centro de Biología Molecular (CSIC-UAM), Universidad Autónoma de Madrid, E-28049 Madrid, Spain[2]; and Unidad de Biofísica (CSIC-UPV/EHU) and Departamento de Bioquímica, Universidad del País Vasco, E-48080 Bilbao, Spain[3]

Received 16 June 2011/Accepted 29 July 2011

Virus infections can result in a variety of cellular injuries, and these often involve the permeabilization of host membranes by viral proteins of the viroporin family. Prototypical viroporin 2B is responsible for the alterations in host cell membrane permeability that take place in enterovirus-infected cells. 2B protein can be localized at the endoplasmic reticulum (ER) and the Golgi complex, inducing membrane remodeling and the blockade of glycoprotein trafficking. These findings suggest that 2B has the potential to integrate into the ER membrane, but specific information regarding its biogenesis and mechanism of membrane insertion is lacking. Here, we report experimental results of *in vitro* translation-glycosylation compatible with the translocon-mediated insertion of the 2B product into the ER membrane as a double-spanning integral membrane protein with an N-/C-terminal cytoplasmic orientation. A similar topology was found when 2B was synthesized in cultured cells. In addition, the *in vitro* translation of several truncated versions of the 2B protein suggests that the two hydrophobic regions cooperate to insert into the ER-derived microsomal membranes.

Virus infections can lead to a variety of cellular injuries, and usually these involve the restructuring of host membrane systems. Viroporins are a group of small virally encoded proteins that interact with cellular membranes to modify permeability and promote the release of viral particles. A typical feature exhibited by viroporins is the presence of at least one membrane-spanning helix anchoring the protein into membranes. After membrane insertion, their oligomerization creates hydrophilic channels or pores (22).

Poliovirus is the enterovirus prototype member of the *Picornaviridae* family. This small, nonenveloped, icosahedral virus possesses a single-stranded 7.5-kb positive-sense RNA genome that encodes a single polyprotein. Polyprotein processing by virus-encoded proteases yields the structural P1 region proteins that encapsidate viral RNA and the nonstructural P2 and P3 region proteins involved in the replication of the viral RNA and membrane permeabilization (2). Nonstructural 2B protein is one of the products generated on processing the P2 region (62). Viroporin 2B has been identified as one of the viral proteins responsible for the alterations in host cell membrane permeability that take place in enterovirus-infected cells. Different 2B proteins expressed in cells have been localized at the endoplasmic reticulum (ER), Golgi complex, and, to a lesser extent, to the plasma and mitochondrial membranes (18, 31, 49, 58). Biochemical and structural data indicate that viroporins form homo-oligomers that create pores in the ER and Golgi complex membranes (1, 16, 17, 30, 59). However, experimental data dealing with the mechanism of the membrane integration of the 2B product are lacking to date.

The poliovirus 2B viroporin protein is hydrophobic overall and rather small (97 amino acids). Hydrophobicity within the viroporin 2B sequence seems to cluster in two main regions (Fig. 1), one predicted to form a cationic amphipathic α-helix located between residues 35 and 55 (HR1) and a second, more hydrophobic, α-helix located between residues 61 and 81 (HR2), as previously suggested (1). Mutations in the amphipathic α-helix or the second hydrophobic region were shown to interfere with the ability of 2B to increase membrane permeability to promote virus release (3, 8, 58) and with viral RNA replication (57), indicating that the soundness of these regions is essential for viral infection. The amphipathic α-helices of several 2B proteins contain three lysine residues at similar positions, and the presence of an aspartic acid residue also is common in the hydrophobic HR2 region (15), suggesting an α-helical hairpin structure.

Two views of the insertion process of α-helical hairpins into the membrane bilayer can be envisioned. One view postulates α-helical hairpin insertion to be a spontaneous process that does not require specific machinery (9, 19). The other supposes a role for the translocon, which is responsible for facilitating the translocation of secreted proteins across the membrane and insertion of membrane proteins into the lipid bilayer (27, 46), allowing *en bloc* α-helical hairpin insertion from a proteinaceous environment into the lipid bilayer (51).

In the present study, we find that viroporin 2B is an integral membrane protein that can be inserted into the ER membrane through the translocon. The *in vitro* translation of model integral membrane protein constructs in the presence of microsomal membranes initially suggested that when expressed separately, only the amphipathic helix (HR1) can span the membrane. However, the *in vitro* translation of truncated versions of the 2B protein carrying appropriate C-terminal re-

* Corresponding author. Mailing address: Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46100 Burjassot, Spain. Phone: 34-96-354 3796. Fax: 34-96-354 4635. E-mail: Ismael.Mingarro@uv.es.
† These authors contributed equally to this work.
▽ Published ahead of print on 10 August 2011.

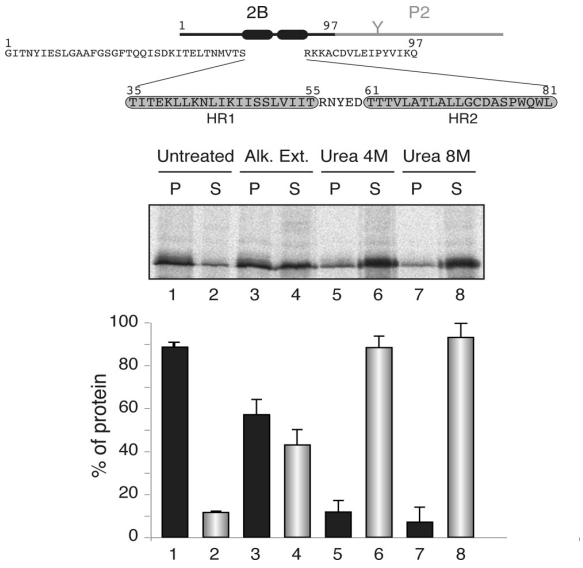


FIG. 1. Membrane association of the chimeric viroporin 2B/P2 protein. At the top is a schematic representation of the poliovirus 2B/P2 chimeric protein (the fused P2 domain is shown in gray). Amino acid residues are shown (HR1 and HR2 are highlighted in gray boxes). The gel in the middle shows the segregation of [^{35}S]Met/Cys-labeled viroporin 2B/P2 fusion protein into membranous and soluble fractions (untreated) and after alkaline wash (Alk. Ext.; sodium carbonate buffer) or urea treatments. P and S denote pellet and supernatant, respectively. In the graph at the bottom, to calculate the percentages of protein the signals present in each pellet and supernatant pair were summed and set to 100%. Data correspond to averages from at least three independent experiments; error bars show standard deviations.

porter glycosylation tags further demonstrated that (i) the N-terminal hydrophobic domain may be stably inserted into the ER-derived microsomal membranes through the translocon, provided that an N_{cyt}/C_{lum} topology is preserved (N_{cy+}, N-terminal end of the transmembrane [TM] segment is oriented toward the cytosol; C_{lum}, C-terminal end of the TM segment is oriented toward the lumen); and (ii) within the complete 2B protein the two hydrophobic regions cooperate to insert into the ER membrane as a helical hairpin with an N-/C-terminal cytoplasmic orientation. In addition, a similar topology was adopted by viroporin 2B expressed in cultured cells under conditions leading to plasma membrane permeabilization.

MATERIALS AND METHODS

Enzymes and chemicals. All enzymes (unless indicated otherwise) as well as plasmid pGEM1, the RiboMAX SP6 RNA polymerase system, and rabbit reticulocyte lysate (a cell-free translation system) were purchased from Promega (Madison, WI). The ER rough microsomes from dog pancreas were purchased from tRNA Probes (College Station, TX). To ensure consistent performance with minimal translational inhibition and background noise, microsomes have been isolated free from contaminating membrane fractions and stripped of endogenous membrane-bound ribosomes and mRNA. The [^{35}S]Met/Cys and ^{14}C-methylated markers were purchased from GE Healthcare. The restriction enzymes and endoglycosidase H (EndoH) were purchased from Roche Molecular Biochemicals. The DNA plasmid, RNA clean-up, and PCR purification kits were from Qiagen (Hilden, Germany). The PCR QuikChange mutagenesis kit was from Stratagene (La Jolla, CA). All oligonucleotides were purchased from Thermo (Ulm, Germany).

Computer-assisted analysis of viroporin 2B sequence. The prediction of TM helices was done using up to 10 of the most common methods available on the Internet: DAS (14) (http://www.sbc.su.se/~miklos/DAS), PHDhtm (47) (http://www.predictprotein.org/), MEMSAT3 (28) (http://bioinf.cs.ucl.ac.uk/psipred/), MEMSAT-SVM (43) (http://bioinf.cs.ucl.ac.uk/psipred/), SOSUI (26) (http://bp.nuap.nagoya-u.ac.jp/sosui/), TMHMM (29) (http://www.cbs.dtu.dk/services/TMHMM), TMPred (http://www.ch.embnet.org/software/TMPRED_form.html), ΔG Prediction Server (24, 25) (http://www.cbr.su.se/DGpred/), SPOCTOPUS (60) (http://octopus.cbr.su.se/), and TopPRED (12) (http://www.sbc.su.se/~erikw/toppred2/).

DNA manipulations. Plasmids encoding full-length 2B sequence (without a stop codon) were constructed by subcloning poliovirus serotype 1 (PV1) Mahoney strain 2B-encoding DNA (kindly provided by E. Wimmer, Stony Brook University) (56) into pGEM1 vector between the NcoI and NdeI restriction sites. This construct contained the P2 domain of the *Escherichia coli* leader peptidase (Lep) fused in frame at the C terminus, as described previously (41). Alternatively, we prepared templates for the *in vitro* transcription of the truncated *2B* mRNA with a 3′ glycosylation tag. The truncated viroporin 2B sequence was prepared by the PCR amplification of a fragment of the pGEM1 plasmid. The 5′ primer was the same for all PCRs and had the sequence 5′-TTCGTCCAACC AAACCGACTC-3′. This primer was situated 210 bases upstream of the *2B* translational start codon; thus, all amplified fragments contained the SP6 transcriptional promoter from pGEM1. The 3′ primers were designed to have approximately the same annealing temperature as the 5′ primer. They contained an optimized glycosylation tag followed by tandem translational stop codons, TAG and TAA, and annealed at specific positions to obtain the desired polypeptide length. PCR amplification comprised a total of 30 cycles with an annealing temperature of 52°C. The amplified DNA products were purified with the Qiagen PCR purification kit according to the manufacturer's protocol and verified on a 1% agarose gel.

In addition, the hydrophobic regions from 2B were introduced into the modified Lep sequence from the pGEM1 plasmid (24, 34) between the SpeI and KpnI sites using two double-stranded oligonucleotides with overlapping overhangs at the ends. The complementary oligonucleotide pairs first were annealed at 85°C for 10 min and then slowly cooled to 30°C, after which the two annealed double-stranded oligonucleotides were mixed, incubated at 65°C for 5 min, cooled slowly to room temperature, and ligated into the vector. The replacements of Lys 46 by Gly, Glu, Gln, and Arg in the LepHR1 construct were done using the QuikChange mutagenesis kit from Stratagene (La Jolla, CA) according to the manufacturer's protocol. All DNA manipulations were confirmed by the sequencing of plasmid DNAs.

Expression *in vitro*. Full-length 2B DNA was amplified from 2B/P2 plasmid using a reverse primer with a stop codon at the end of the 2B sequence (2B-derived expressions), or the DNA derived from the pGEM1 plasmid was transcribed directly (2B/P2 construct). Alternatively, viroporin 2B was amplified fused to the first 50 amino acids from P2 using a reverse primer with tandem stop codons at the 3′ end (2B/50P2). The transcription of the DNA derived from the pGEM1 plasmid was done as described previously (61). Briefly, the transcription mixture was incubated at 37°C for 2 h. The mRNA products were purified with a Qiagen RNeasy clean-up kit and verified on a 1% agarose gel.

In vitro translation of *in vitro*-transcribed mRNA was performed in the presence of reticulocyte lysate, [^{35}S]Met/Cys, and dog pancreas microsomes as described previously (21, 61). Lep constructs with HR-tested segments were transcribed and translated as previously reported (33, 34). For the posttranslational membrane insertion experiments, 2B-derived mRNAs were translated (37°C for 1 h) in the absence of rough microsomal membranes (RMs). Translation then was inhibited with cycloheximide (10 min at 26°C; 2 mg/ml final concentration), after which RMs were added and incubated for an additional hour at 37°C. In all cases, after translation membranes were collected by ultracentrifugation and analyzed by sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Finally, the gels were visualized on a Fuji FLA3000 phosphorimager with ImageGauge software.

For EndoH treatment, the translation mixture was diluted in 4 volumes of 70 mM sodium citrate (pH 5.6) and centrifuged (100,000 × *g* for 20 min at 4°C). The pellet was resuspended in 50 μl of sodium citrate buffer with 0.5% SDS and 1% β-mercaptoethanol, boiled for 5 min, and incubated for 1 h at 37°C with 0.1 mU of EndoH. The samples were analyzed by SDS-PAGE.

For the proteinase K protection assay, the translation mixture was supplemented with 1 μl of 50 mM $CaCl_2$ and 1 μl of proteinase K (4 mg/ml) and then digested for 40 min on ice. The reaction was stopped by adding 1 mM phenylmethylsulfonyl fluoride (PMSF) before SDS-PAGE analysis.

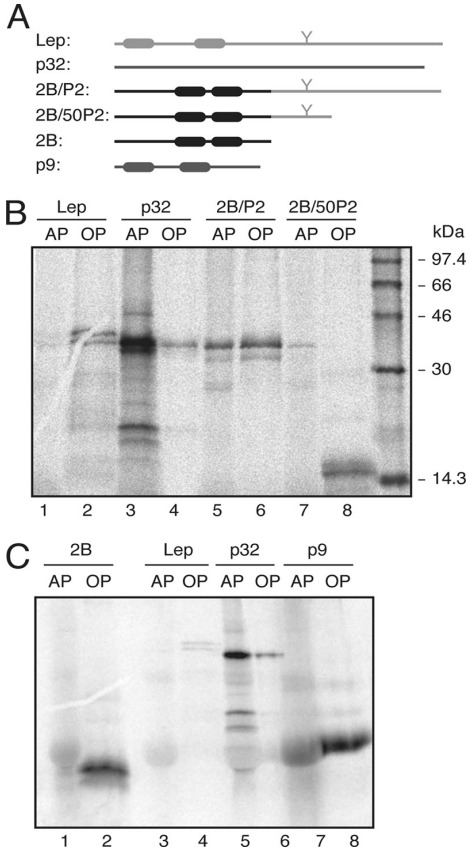


FIG. 2. Triton X-114 partition of viroporin 2B and 2B-derived proteins. (A) Structural organization of the proteins used in the Triton X-114 partition experiments. (B) SDS-PAGE (12% polyacrylamide) analysis after Triton X-114 treatment of 2B/P2 (~38 kDa) and 2B/50P2 (~16 kDa) proteins. Integral membrane protein Lep (~37 kDa; the nonglycosylated form) and peripheral PNRSV movement protein (~32 kDa) were processed in parallel as control samples. (C) Phase separation of viroporin 2B. As a control for small integral membrane protein, Turnip crinkle virus p9 movement protein (~9 kDa) was included. The gels used contained 20% polyacrylamide. AP and OP refer to aqueous and organic phases, respectively.

Membrane sedimentation, alkaline wash, and urea treatments. The translation mixture was diluted in 8 volumes of buffer A (35 mM Tris-HCl at pH 7.4 and 140 mM NaCl) for the membrane sedimentation, 4 volumes of buffer A supplemented with 100 mM Na_2CO_3 (pH 11.5) for the alkaline wash, and 4 or 8 M urea for urea treatments. The samples were incubated on ice for 30 min and clarified by centrifugation (10,000 × *g* for 20 min). Membranes were collected by the ultracentrifugation (100,000 × *g* for 20 min at 4°C) of the supernatant onto 50-μl sucrose cushions. Pellets (P) and supernatants (S) of the ultracentrifugation were analyzed by SDS-PAGE.

Phase separation in Triton X-114 solution. The phase separation of integral membrane proteins using the detergent Triton X-114 was performed as described previously (10, 41). Triton X-114 (1%, vol/vol) was added to a translation mixture that previously had been diluted with 180 μl of phosphate-buffered saline (PBS). After mixing, the samples were incubated at 0°C for 1 h and overlaid onto 300 μl of PBS supplemented with 6% (wt/vol) sucrose and 1% (vol/vol) Triton X-114. After 10 min at 30°C, an organic droplet was obtained by centrifugation for 3 min at 1,500 × *g*. The resulting aqueous upper phase (AP; 200 μl) was collected, and the organic droplet at the bottom of the tube was diluted with PBS (organic phase [OP]). Both OP and AP were supplemented with sample buffer and boiled for 10 min prior to 12% (Fig. 2B) or 20% (Fig. 2C) SDS-PAGE analysis.

Transfection assay. Baby hamster kidney (BHK) cells that stably express the T7 RNA polymerase (clone BSR-T7/5), designated BHKT7 (11), were used. Cells were grown at 37°C in Dulbecco's modified Eagle's medium (DMEM) supplemented with 5 or 10% fetal calf serum (FCS) and nonessential amino acids. BHKT7 cells additionally were treated with Geneticin G418 (Sigma) on every third passage at a final concentration of 2 mg/ml. Cells were transfected with 1 μg of plasmid pTM1-2B (3, 4) or the different constructs plus 2 μl of Lipofectamine per well in Opti-mem medium (Invitrogen) for 2 h at 37°C. After 2 h, Lipofectamine was removed and the cells were supplemented with fresh medium containing 5% FCS.

RESULTS

Viroporin 2B is an integral membrane protein. Viroporin 2B amino acid sequence (Fig. 1) has been parsed to test the performance of several commonly used algorithms for predicting the topology of integral membrane proteins. As shown in Table 1, the predicted outcome showed great variability according to the methods used, likely due to the limited hydrophobicity of the membrane-associating regions of 2B. The membrane association properties of the full-length viroporin 2B were studied using an *in vitro* system that closely mimics the *in vivo* situation, in which cytosolic and membrane fractions of *in vitro*-translated [^{35}S]Met/Cys-labeled 2B/P2 fusions in the presence of ER-derived microsomes were collected and analyzed by SDS-PAGE and autoradiography. In this system, the microsomes provide all of the membrane insertion and glycosylation components (i.e., the translocon machinery and the oligosaccharyltransferase enzyme). The reporter P2 domain is the extramembrane C-terminal domain from the bacterial leader peptidase (Lep) that carries an N-glycosylation site extensively used to report membrane translocation (Fig. 1, top). The viroporin 2B/P2 chimera was recovered mainly from the 100,000 × *g* pellet fraction (Fig. 1, untreated lanes) after the centrifugation of the microsome-containing translation reaction mixture, indicating that it could be either a membrane-associated protein or a lumenally translocated protein. The absence of glycosylation suggested that the chimeric protein was not translocated into the lumen of the microsomes. Nevertheless, to differentiate between these possibilities the translation reaction mixtures were washed with sodium carbonate (pH 11.5), which renders microsomes into membranous sheets, releasing the soluble luminal proteins (35, 41). As shown in Fig. 1, the 2B/P2 fusion appeared to be preferentially associ-

TABLE 1. Computer analysis of viroporin 2B amino acid sequence

Algorithm	Membrane protein	No. of TM segments (starting aa/ending aa)
DAS	Yes	2 (45/53–64/71)
PHDhtm	Yes	2 (41/57–64/72)
MEMSAT3	Yes	1 (44/63)
MEMSAT-SVM	Yes	1 (40/55)
SOSUI	No	0
TMHMM	No	0
TMPred	Yes	1 (58/78)
ΔG Prediction Server	Yes	1 (61/82)
SPOCTOPUS	Yes	1 (48/73)
TopPRED	Yes	1 (38/58)

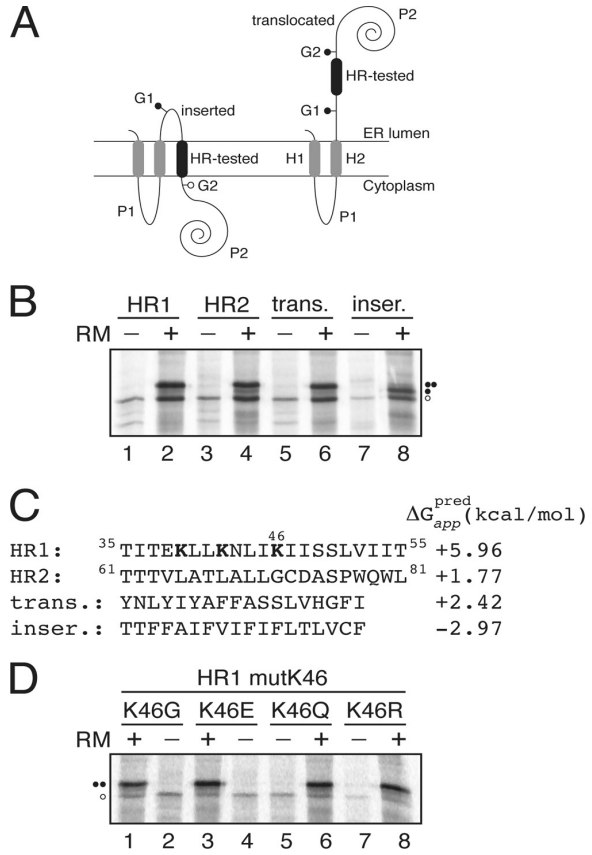


FIG. 3. Insertion of viroporin 2B hydrophobic regions 1 (HR1) and 2 (HR2) into microsomal membranes using Lep as a model protein. (A) Schematic representation of the leader peptidase (Lep) construct used to report insertion into the ER membrane of 2B HR1 and HR2. The HR under study is inserted into the P2 domain of Lep flanked by two artificial glycosylation acceptor sites (G1 and G2). The recognition of the HR by the translocon machinery as a TM domain locates only G1 in the luminal side of the ER membrane, preventing G2 glycosylation. The Lep chimera will be doubly glycosylated when the HR being tested is translocated into the lumen of the microsomes. (B) *In vitro* translation in the presence of membranes of the different Lep constructs. Constructs containing HR1 (residues 35 to 55; lanes 1 and 2) and HR2 (residues 61 to 81; lanes 3 and 4) were transcribed and translated in the presence of membranes. Control HRs were used to verify sequence translocation (trans.; lanes 5 and 6) and membrane integration (inser.; lanes 7 and 8). Bands of nonglycosylated protein are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. (C) The HR sequence in each construct is shown together with the predicted ΔG apparent value, which was estimated using the ΔG prediction algorithm available on the Internet (http://dgpred.cbr.su.se/). Lysine residues in HR1 are shown in boldface. (D) *In vitro* translation of HR1-derived mutants at lysine 46.

ated (approximately 58%; lanes 3 and 4) with the membranous pellet fraction, suggesting a tight association with membranes. Further treatment with 4 M urea demonstrated that approximately 88% of the protein was in the supernatant fraction (Fig. 1, lanes 5 and 6). More than 93% of the protein was extracted from the supernatant fraction by 8 M urea (Fig. 1, lanes 7 and 8), suggesting that 2B/P2 is released from the membrane environment when the secondary and tertiary structures of the protein are lost.

The translation reaction mixtures also were treated with Triton X-114, a nonionic detergent that forms a separate organic phase into which membrane lipids and hydrophobic proteins are segregated from the aqueous phase, which contains nonintegral membrane proteins (10, 35). The 2B/P2 fusion protein (Fig. 2A) was detected in both the aqueous and organic phases (Fig. 2B, lanes 5 and 6), while a reduction of the P2 domain to its 50 N-terminal residues (Fig. 2A) led to the organic-phase detection of the chimera (Fig. 2B, lanes 7 and 8). Control analyses of Lep and Prune necrotic ring-spot virus (PNRSV) p32 movement protein showed, as previously demonstrated (35), organic- and aqueous-phase detections, respectively. Finally, viroporin 2B translated in the absence of fused domains was detected only in the organic phase (Fig. 2C, lanes 1 and 2), indicating that viroporin 2B is an integral membrane protein.

Insertion of the viroporin 2B hydrophobic regions into biological membranes. We assayed the membrane insertion capacities of the viroporin 2B hydrophobic regions using an *in vitro* experimental system (24), which accurately reports the integration of transmembrane (TM) helices into microsomal membranes. This system uses ER-derived microsomal membranes and provides a sensitive way to detect the insertion or translocation of hydrophobic regions through the Sec translocon (25). An obvious advantage of this system is that the insertion assays are performed in the context of a biological membrane. The system is based on the cotranslational glycosylation performed by the oligosaccharyltransferase (OST) enzyme. OST adds sugar residues to an NX(S/T) consensus sequence (53), with X being any amino acid except proline (52), after the protein emerges from the translocon channel. The glycosylation of a protein region translated *in vitro* in the presence of microsomal membranes therefore indicates the exposure of this region to the OST active site on the luminal side of the ER membrane. In our first experimental assay (Fig. 3), a segment to be assayed (HR tested) is engineered into the luminal P2 domain of the integral membrane protein Lep from *E. coli*, where it is flanked by two acceptor sites (G1 and G2) for *N*-linked glycosylation (Fig. 3A). Both engineered glycosylation sites are to be used as membrane insertion reporters. The rationale behind using two glycosylation sites is that G1 will always be glycosylated due to its native luminal localization, while G2 will be glycosylated only on the translocation of the tested TM region through the microsomal membrane. Thus, single glycosylation indicates a correct TM integration (Fig. 3A, left), whereas double glycosylation reports the noninteration capability of the tested HR segment (Fig. 3A, right). The single glycosylation of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep expressed in the absence of microsomes, and the mass is around 5 kDa in the case of double glycosylation.

The translation of the chimeric constructs harboring the predicted viroporin 2B hydrophobic regions as HR-tested segments resulted mainly in double-glycosylated forms (Fig. 3B,

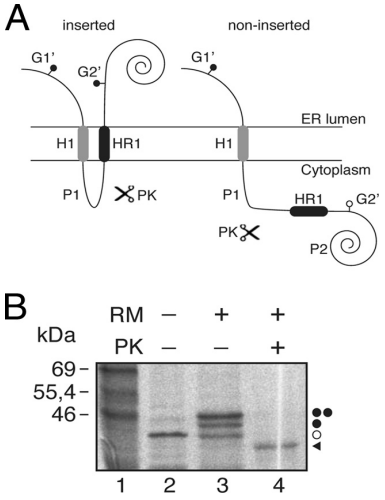


FIG. 4. Insertion of HR1 into microsomal membranes using the Lep′ construct. (A) Schematic representation of the Lep-derived construct (Lep′). In this Lep′ construct (~40 kDa; the nonglycosylated form), HR1 replaces the H2 domain from Lep. The glycosylation acceptor site (G2′) located in the beginning of the P2 domain will be modified only if HR1 inserts into the membrane, while the G1′ site, embedded in an extended N-terminal sequence of 24 amino acids, is always glycosylated. (B) *In vitro* translation in the presence of membranes. Bands of nonglycosylated protein are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. The protected glycosylated HR1/P2 fragment (~36.5 kDa) is indicated by a black triangle.

lanes 2 and 4), which is consistent with the translocation of these regions into the lumen of the ER. Control constructs with previously tested (37) translocation and integration sequences are shown in lanes 6 and 8, and they disclose the expected double and single glycosylation patterns, respectively (Fig. 3B). The permeabilization induced by an overlapping peptide library that spanned the complete viroporin 2B sequence mapped the cell plasma membrane-porating activity to the partially amphipathic HR1 domain (32). This region contains three lysine residues that would preclude TM disposition (Fig. 3C), especially in the case of lysine 46, since it would be located roughly in the center of the hydrophobic core of the lipid bilayer. Nevertheless, the replacement of this residue by glycine, glutamic acid, glutamine, or arginine renders glycosylation patterns consistent with the translocation of the hydrophobic region into the ER lumen (Fig. 3D). Taken together, these results suggest that an isolated HR1 segment does not span ER-derived membranes in an N_{lum}/C_{cyt} orientation.

Since native 2B does not have a cleavable signal sequence, it seems likely that HR1 acts as a signal-anchor sequence having an N_{cyt}/C_{lum} orientation in the membrane (36). To test HR1 insertion in a reverse orientation, another Lep construct (Lep′) was used. In this Lep′ construct, HR1 replaces the second TM segment (H2) from Lep (Fig. 4A). The glycosylation acceptor site (G2′) located in the beginning of the P2 domain will be modified only if the HR1 segment inserts into the membrane, while the G1′ site, embedded in an extended N-terminal sequence of 24 amino acid residues, is always glycosylated. We found that HR1 significantly inserts into the membrane (up to 60% of the molecules) with the appropriate topology (Fig. 4B). The nature of the cytosolic/luminal domains was further examined by proteinase K (PK) digestions. Treatment with PK degrades domains of membrane proteins that protrude into the cytosol, but membrane-embedded or luminally exposed domains are protected. The addition of PK to a Lep′HR1 translation mixture (Fig. 4B, lane 4) rendered a protected, glycosylated HR1-P2 fragment, suggesting the proper insertion of HR1 sequence with an N_{cyt}/C_{lum} orientation.

Viroporin 2B integrates into the ER membrane through the translocon with an N-terminal/C-terminal cytoplasmic orientation. The microsomal *in vitro* system closely mimics the conditions of *in vivo* membrane protein assembly into the ER membrane. HR1 is properly recognized by the translocon as a TM segment out of its native context (Fig. 4). However, the presence of fused domains can influence its membrane insertion capacity. Hence, we next sought to investigate whether HR1 also could direct the integration into the ER membrane of the native 2B sequence (i.e., in the absence of nonviral fused domains) through the translocon.

Because N-glycosylation acceptor sites are absent from the viroporin 2B sequence, several modifications were prepared to determine the TM disposition of different 2B-derived proteins. First, to gain topological information, an N-glycosylation acceptor site was engineered at the hydrophilic N-terminal region of the 2B sequence by mutating glutamine 20 to an acceptor asparagine (…^{20}NIS…; construct 2BNtGlyc). Second, we added a C-terminal N-glycosylation tag (CtGlyc; **NST-**MMM [the glycosylation sequon is in boldface]) that has been proven to be efficiently glycosylated (6). The first 60 residues of the viroporin 2B carrying an N-terminal glycosylation site were translated using C-terminal tags (Fig. 5A) either with an N-glycosylation acceptor site as a C-terminal tag (2BNt/Ct; Fig. 5B, lanes 1, 3, and 5) or with a nonacceptor site (2BNt/CtØ; Fig. 5B, lanes 2, 4, and 6). The lack of glycosylation at the N-terminal engineered acceptor site together with the efficient glycosylation observed only when using the C-terminal acceptor site (Fig. 5B, lane 3), as proven after EndoH treatment (Fig. 5B, lane 5), strongly indicates that HR1 in the viroporin context is acting as a noncleavable signal sequence and is properly recognized by the translocon machinery to be inserted into the membrane with its N terminus facing the cytoplasm (N_{cyt}/C_{lum} topology). In addition, by blocking protein synthesis after 2BNt/Ct has been translated in the absence of membranes, we confirmed that the truncated version of 2B (2BNt/Ct) needs to be cotranslationally inserted into the ER membrane. As shown in Fig. 5C, 2BNt/Ct was glycosylated when microsomal membranes were added to the translation mixture cotranslationally. In contrast, when microsomal membranes were included posttranslationally (i.e., after translation had been inhibited by cycloheximide), the C-terminal acceptor site was not glycosylated (Fig. 5C, lane 3), thereby emphasizing that truncated 2B is integrated cotranslationally through the ER translocon.

Because of the low hydrophobicity and the relatively poor insertion propensity found in the Lep system for HR2 (Fig. 3), it is predicted that the 2B C-terminal region will be translocated into the ER lumen. However, 2B hydrophobic regions also are responsible for the membrane anchoring of the 2BC

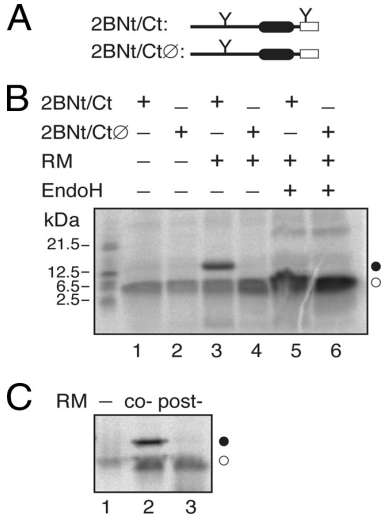


FIG. 5. Membrane insertion of 60-mer viroporin 2B. (A) Structural organization of the 2B 60-mer truncated construct. (B) *In vitro* translations were performed in the presence (+) or in the absence of C-terminal glycosylation tag, RMs, and EndoH as indicated. (C) The 2BNt/Ct construct was translated in either the absence (lanes 1 and 3) or the presence (lane 2) of RMs. Lane 2, cotranslationally added microsomes. Lane 3, RMs were added posttranslationally (after 1 h of translation and 10 min of cycloheximide treatment; Post-), and incubation was continued for another 1 h.

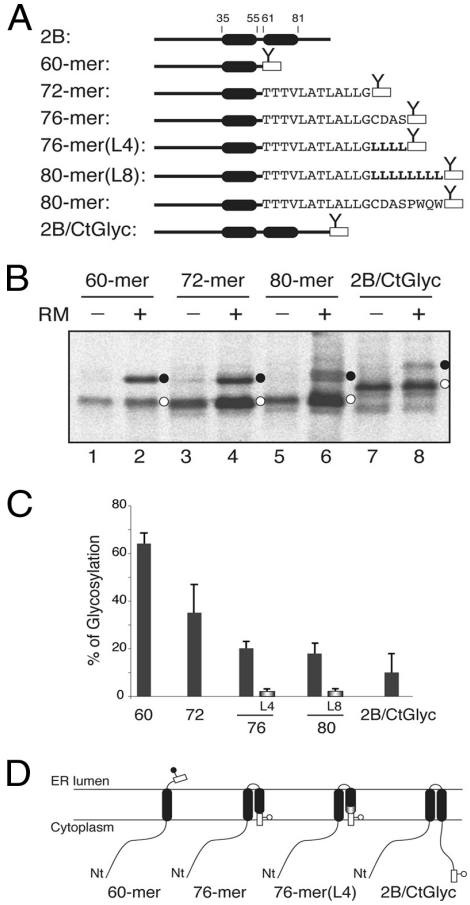


FIG. 6. Effect of HR2 on 2B 60-mer insertion and topology. (A) Structural organization of full-length and truncated viroporin 2B constructs. (B) *In vitro* translation of truncated viroporin 2B 60-mer, 72-mer, 80-mer, and full-length (2B/CtGlyc) constructs in which a fused C-terminal N-glycosylation tag (rectangle) provides a simple readout for topology determination. The presence of RMs and nonglycosylated and glycosylated proteins (empty and black dots, respectively) is indicated. (C) *In vitro* glycosylation of truncated viroporin 2B-derived proteins. The level of glycosylation is quantified from SDS-PAGE gels by measuring the fraction of glycosylated (f_g) versus glycosylated-plus-nonglycosylated (f_{ng}) molecules, using the equation $p = f_g/f_g + f_{ng}$. Data correspond to averages from at least four independent experiments, and error bars show standard deviations. (D) Topological models for 2B constructs. Nt, N terminus.

precursor, which performs its enzymatic and RNA binding activities in the cytosolic compartment. Thus, we speculated that some type of helix-helix interaction stabilizes the insertion of HR2 to keep an N_{cyt}/C_{cyt} 2B topology. It has been shown recently that in some cases the insertion of poorly hydrophobic regions depends on the presence of neighboring loops and/or TM segments (20), especially in the case of the preceding TM segment (23, 55). In our attempts to unravel the disposition of 2B in biological membranes, we focused on the insertion of truncated C-terminal reporter tag fusions. In Fig. 6 we show a series of experiments where HR1 and various lengths of downstream sequence were translated as truncated proteins with an N-glycosylation C-terminal tag (Fig. 6A). A mutant polypeptide truncated at the end of the hydrophilic loop between HR1 and HR2 (60-mer) is highly glycosylated (~65%; Fig. 6B, lane 2), indicating that, similarly to what is shown in Fig. 5B, the C-terminal glycosylation tag has been translocated into the lumen of the ER, and thus HR1 is integrated into the ER membrane in a N_{cyt}/C_{lum} orientation in this construct (Fig. 6D, left). The percentage of glycosylated truncated proteins is reported in Fig. 6C. Extending the 2B sequence to include roughly half of HR2 has a significant effect on this pattern (72-mer; Fig. 6B and 6C), suggesting some tendency of these truncated molecules to insert the C-terminal tag into the membrane. Moreover, extending the 2B sequence four residues (roughly one helical turn; 76-mer) (Fig. 6C) substantially diminished glycosylation (~21%). The glycosylation level for the truncated protein shown in Fig. 6C cannot be explained by an increased hydrophobicity of the added amino acids, since the total free energy predicted (ΔG^{pred}) for the ^{73}CDAS^{76} sequence is 4.31 kcal/mol, where a positive value is indicative of extramembrane disposition (the calculation of ΔG^{pred} was carried out using the scale of Hessa and collaborators [24]). Interestingly, the addition of four leucine residues (ΔG^{pred} = −2.2 kcal/mol) instead of the ^{73}CDAS^{76} sequence [76-mer(L4)] strongly precludes glycosylation (<3%) (Fig. 6C), demonstrating the clear hydrophobic effect of the leucine residues in this construct. This was further corroborated by extending the protein to include up to eight leucine residues

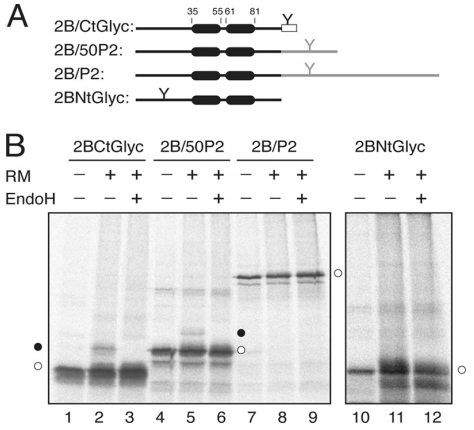


FIG. 7. Insertion and topology of full-length viroporin 2B protein. (A) The structural organization of the 2B-derived constructs is shown at the top. The N-glycosylation site is highlighted by a Y-shaped symbol both when inserted in the protein sequence and when added as a C-terminal reporter tag (rectangles). (B) *In vitro* translation was performed in the presence (+) and in the absence (−) of RMs and EndoH as indicated. Nonglycosylated and singly glycosylated proteins are indicated by empty and black dots, respectively.

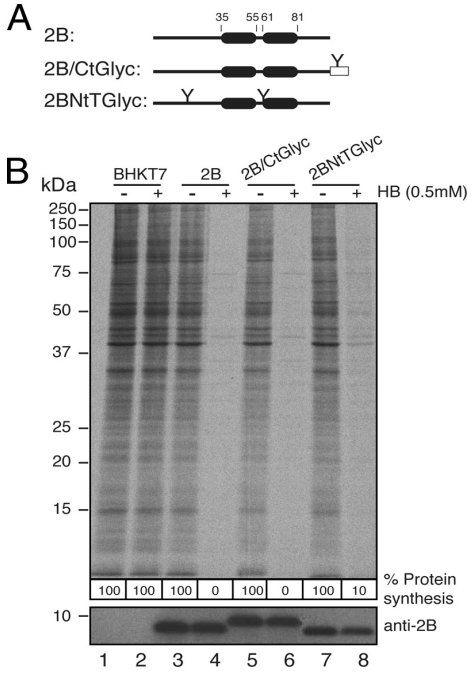


FIG. 8. Permeabilization activity of 2B-derived constructs in BHKT7 cells. (A) Structural organization of the transfected 2B variants. (B) BHKT7 cells were transfected with pTM1-2Bwt or with pTM1-2B variants. At 3 h posttransfection, cells were pretreated with hygromycin B (HB) at 0.5 mM for 15 min and then labeled with [^{35}S]Met/Cys for 45 min in the presence of the inhibitor. After labeling, the proteins were analyzed by SDS-PAGE (17.5%) followed by fluorography and autoradiography (top). The synthesis of 2B protein was detected by Western blotting using specific rabbit polyclonal antibodies (bottom). The numbers below each lane represent the percentage of protein synthesis in the presence of hygromycin B as calculated by densitometric scanning.

[80-mer(L8)] (Fig. 6C). Finally, both the mutant truncated at the end of HR2 (80-mer) and the 2B full-length construct (2B/CtGlyc) had little effect on this pattern (~17 and ~10% glycosylation, respectively), indicating that the C terminus of the majority of these tagged proteins is cytosolic, and thus the HR2 sequence included in these constructs is integrated into the membrane in an N_{lum}/C_{cyt} orientation (Fig. 6D, right).

To confirm that the same topology is adopted by the full-length 2B protein, several 2B-derived constructs were prepared and their membrane disposition experimentally determined. The translation of the C-terminal-tagged 2B protein (2B/CtGlyc) (Fig. 7A) in the presence of RMs resulted in glycosylation in <8% of the molecules, as demonstrated by EndoH (a glycan-removing enzyme) treatment (Fig. 7B, lanes 1 to 3). The addition of the first 50 residues of the Lep P2 domain (2B/50P2) (Fig. 7A), which contains an N-glycosylation acceptor site as a topological reporter, yielded glycosylation in <5% of the viroporin 2B-derived molecules (Fig. 7B, lanes 4 to 6). Furthermore, when full-length Lep P2 domain was used as a reporter domain, the chimera was not glycosylated at all (2B/P2) (Fig. 7B, lanes 7 to 9). Finally, the *in vitro* translation of a construct harboring an N-terminal glycosylation acceptor site (2BNtGlyc, Fig. 7A) in the presence of RM only resulted in unmodified molecules (Fig. 7B, lanes 10 to 12). Taken together, these results suggest a preferential N-/C-terminal cytoplasmic orientation for viroporin 2B when expressed in the presence of ER-derived microsomal membranes.

Viroporin 2B topology in mammalian cells. To further assess the topology adopted by functional viroporin 2B in membranes, 2B variants containing designed N-glycosylation sites at different positions were expressed in cultured cells, and their plasma membrane-permeabilizing capacities were assessed (Fig. 8). For this purpose, the 2B variants were cloned in pTM1 vector and transfected in BHK cells that stably express the T7 RNA polymerase. These cells posses the machinery required for synthesizing the virion components and even to assemble infectious particles. As expected, 2B expression permeabilized BHK cells to the antibiotic hygromycin B (31) (Fig. 8B, lane 4). Although the synthesis of unmodified 2B is not detectable by radioactive labeling in the permeabilized cells (Fig. 8B, top), it can be detected by Western blotting using a specific 2B antibody (Fig. 8B, bottom). No glycosylation of 2B was observed when this protein bears the N-glycosylation site at the amino terminus, in the turn, or at the carboxy terminus of this viroporin (Fig. 8B, lanes 5 to 8). In addition, these two viroporin 2B variants retain their capacity to permeabilize cells to hygromycin B, suggesting that they are located at the membrane and exhibit the ability to alter membrane permeability. It should be noted that in the case of the N-glycosylation site located at the turn, although the location is luminal the absence of glycosylation is due to its proximity to the TM domains as previously reported (42, 44). In conclusion, the *in vitro* and *in vivo* assays consistently indicate that both the N and

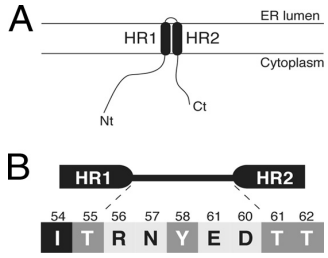


FIG. 9. (A) Topological model for 2B association with membranes. (B) Turn-inducing propensity at the interhelical region of the viroporin 2B helical hairpin according to the scale of Monné and coworkers (39). The five amino acid residues interconnecting HR1 and HR2 are shown flanked by the putative last two residues from the HR1 helix and the putative first two residues from HR2. All residues connecting HR1 and HR2 are turn inducers (normalized turn potential, >1). The residues are highlighted according to their turn potential: black for a potential lower than 1 (Ile54), dark gray for a potential between 1 and 2 (Thr55, Tyr58, and Thr61), and light gray for a potential above 2 (Arg56, Asn57, Glu59, and Asp60).

C termini of viroporin 2B face the cytoplasm, as displayed in Fig. 9A.

DISCUSSION

Viroporins are a group of proteins responsible for alterations in the permeability of cellular membranes during virus infection, favoring the release of viral particles from infected cells (reviewed in reference 22). The molecular mechanisms by which viroporins insert into cell membranes remain largely unknown. In this study, we demonstrate that poliovirus viroporin 2B is a double-spanning integral membrane protein that can be inserted into the ER membrane through the translocon machinery.

Computer-assisted membrane protein topology prediction is a useful starting point for experimental studies of membrane proteins. We have used 10 popular prediction methods and found large discrepancies between their predictions. Two of the algorithms failed to predict 2B as an integral membrane protein, and two of them assigned two TM segments for the protein. It should be noted that the reliability of a topology prediction can be estimated by the number of prediction methods that agree. Since six of the algorithms predicted 2B as a membrane protein with only one TM segment, these results clearly highlight that the presence of helical hairpin structures, which was not detected even by the methods predicting reentrant loops (MEMSAT-SVM and SPOCTOPUS), may be missed by current predictive methods, as previously suggested for a different TM helical hairpin (40).

The membrane association of 2B/P2 fusion protein was resistant to alkaline extraction. Since this treatment disrupts microsomal membranes and releases any soluble luminal protein, this result indicates that the fusion protein is not translocated to the lumen of the microsomes. Urea treatments solubilized our fusion protein (Fig. 1), indicating that secondary and tertiary structures in 2B play an important role in 2B insertion. The latter results contrast with previous work that showed that coxsackievirus 2B/enhanced green fluorescent protein fusions were resistant to urea extraction (18). This discrepancy could be derived either from the differences found in the amino acid sequence of both 2B proteins or from the use of different fusion proteins in both cases. In fact, a significant influence of the P2 domain can be observed in our Triton X-114 partition experiments. Fusions containing the full P2 domain partition significantly into the aqueous phase, whereas the addition of the 50 N-terminal residues from this domain promoted the partitioning of the shorter chimera into the organic phase. These results clearly demonstrated integral membrane protein behavior (compare lanes 5 and 6 with lanes 7 and 8 in Fig. 2B), as corroborated by partition experiments using full-length 2B (Fig. 2C, lanes 1 and 2).

By challenging the hydrophobic regions of 2B in a model protein construct (Lep), we demonstrate first that these regions do not integrate as TM segments in the presence of ER-derived membranes when expressed separately (Fig. 3). It should be mentioned that, in these Lep-derived constructs, the HR1 segment is forced to insert into the membrane with an N_{lum}/C_{cyt} topology, and this topological effect can prevent the proper TM disposition of this region. In fact, using a Lep-derived variant (Lep′), we demonstrated the TM disposition of HR1 when expressed with an N_{cyt}/C_{lum} orientation (Fig. 4). The glycosylation data obtained in the context of the parental 2B sequence using engineered and truncated proteins provided compelling evidence that HR1-HR2 also may integrate cotranslationally into the membrane in the absence of fused domains. Hence, a truncated 2B protein containing HR1 inserted efficiently into ER microsomal membranes adopting an N_{cyt}/C_{lum} topology (Fig. 5), and the addition of HR2 residues to this construct resulted in the cytoplasmic reorientation of the C terminus (Fig. 6). Furthermore, the lack of glycosylation at N-glycosylation acceptor sites engineered at different positions, both in an *in vitro* microsomal system (Fig. 7) and in cultured cells (Fig. 8), suggests that both the N and C termini of viroporin 2B protein reside on the cytosolic side of the membrane. In the context of the P2 polyprotein, the membrane topology found in the present work leaves the protease cleavage sites of P2 facing the cytosol, which is suitable for polypeptide processing by viral proteases.

Taken together, these data support the capacity of the HR domains to act as interacting TM segments in their natural contexts. In this sense, we have described previously the integration into the ER membrane of two closely spaced membrane-spanning segments of viral origin, where both TM segments of the nascent protein bind to one or more translocon proteins and are held until the termination of translation, whereupon they are released laterally as a helical hairpin into the lipid phase (50, 51). More recently, this mechanism of partition into the membrane as a pair of helices has been observed by others using a nonviral membrane protein (13). Thus, the retention of a first cationic amphipathic segment (HR1) at the ER translocon to generate a helical hairpin might facilitate partitioning into the lipid phase by shielding the polar amino acids that could compromise effective membrane integration (36).

These findings provide important new insights into the molecular architecture and the molecular mechanism of 2B integration into the membrane. The synergic effect found for pore formation between HR1 and HR2 in a previous peptide-based

analysis, which indicated that both HRs cooperate in membrane permeabilization (48), suggests that HR1 and HR2 interact with each other to form a helix-turn-helix (helical hairpin) motif that traverses the lipid bilayer. An additional source of stability of this motif can be the turn between the two helices. It has been shown that charged and polar residues (plus prolines and glycines) display turn induction in a TM polyleucine stretch (38); in our case, we do not know exactly which amino acid residues form the turn in the membrane-bound viroporin, although in all likelihood the turn occurs between the highly hydrophilic residues 55 and 61. Figure 9B shows this region of viroporin 2B, where the residues are highlighted according to their turn-inducing propensities (38). All residues present in this region are strongly turn inducing (normalized turn potential, >1). Among them, four (Arg56, Asn57, Glu59, and Asp60) have a high turn potential (>2). In essence, a great concentration of turn-promoting residues is found in the region connecting HR1 and HR2. Thus, in the membrane-bound form, we can expect the turn of viroporin 2B to be centered on ^{56}RNYED^{60}. Interestingly, previous mutagenesis studies using the CBV 2B protein showed that the negatively charged residues found in this short hydrophilic turn between HR1 and HR2 are indeed important for the membrane-active character of 2B protein (16). Moreover, recent molecular dynamic simulations of the poliovirus 2B channel/pore-forming regions suggested that Glu59 and Asp60 are involved in the helical hairpin formation (45). In any case, it seems clear that the turn may play a significant role in the stability and integration of the membrane-bound 2B protein. In addition, the topology observed in the present work agrees with previous data obtained with different fusion proteins (18). Furthermore, the localization of the C terminus at the cytosolic side of the membrane is consistent with the need for the proteolytic liberation of the 2B protein from the precursor 2BC polyprotein by a cytosolic viral protease cleavage, which is accomplished by 3C^{pro} (54). This could occur after the membrane insertion of 2BC or even the entire P2 precursor (2ABC).

Our findings further suggest a physiological role for translocon-mediated 2B integration into the ER membrane. Our combined analysis predicts a marginal propensity for 2B polypeptide to insert into membranes (Table 1 and Fig. 3D). On the other hand, upon viral entry, initially synthesized 2B or 2BC proteins will remain diluted in the cytosol of the infected cell. Marginal hydrophobicity together with low concentration are predicted to reduce the probability of the spontaneous insertion of 2B and 2BC into their primary target organelle: the Golgi complex (18, 49). Thus, we speculate that cotranslational insertion into the ER membrane is a particularly relevant phenomenon at the initial stages of the infectious cycle, during which both the viral mRNA levels and the concentration of the translated viral proteins are predictably low. Under those conditions, the cellular protein biosynthesis and vesicular transport machineries remain functional, and 2B and its 2BC precursor likely are synthesized as additional cell membrane proteins. From the ER, these proteins may reach the Golgi compartment membrane to fulfill regulatory and/or signaling functions that result in the disruption of Ca^{2+} homeostasis (5, 15) and vesicular transport inhibition (7, 15). At later stages, the canonical ER-Golgi protein-trafficking pathway no longer is functional and/or required for viral replication. The massive proliferation of cell endomembranes (viroplasm) and the high levels of viral protein synthesis result in higher effective concentrations of these components inside the infected cell system. Under those conditions, it is predicted that the spontaneous insertion into the membrane of a significant amount of synthesized 2B and 2BC will ensue.

In summary, viroporin 2B may use common structural arrangements to integrate into the ER membrane through the translocon, at least during the initial stages of the viral replicative cycle. The development of *in vitro* assays designed to dissect the membrane integration process will lead to a better understanding of the membrane permeabilization mechanisms that act during enterovirus infection.

ACKNOWLEDGMENTS

We thank Arne Elofsson (Stockholm University) for helpful comments on the manuscript.

This work was supported by grants BFU2009-08401 (to I.M.), BFU2009-07352 (to L.C.), and BIO2008-00772 (to J.L.N.) from the Spanish Ministry of Science and Innovation (MICINN, ERDF; supported by the European Union), as well as by ACOMP/2011/025 from the Generalitat Valenciana to I.M. M.B.-P. was the recipient of an FPU fellowship from the Spanish Ministry of Education.

REFERENCES

1. **Agirre, A., A. Barco, L. Carrasco, and J. L. Nieva.** 2002. Viroporin-mediated membrane permeabilization. Pore formation by nonstructural poliovirus 2B protein. J. Biol. Chem. **277:**40434–40441.
2. **Agol, V.** 2002. Picornavirus genome: an overview, p. 127–148. *In* B. L. Semler and E. Wimmer (ed.), Molecular biology of picornaviruses. ASM Press, Washington, DC.
3. **Aldabe, R., A. Barco, and L. Carrasco.** 1996. Membrane permeabilization by poliovirus proteins 2B and 2BC. J. Biol. Chem. **271:**23134–23137.
4. **Aldabe, R., and L. Carrasco.** 1995. Induction of membrane proliferation by poliovirus proteins 2C and 2BC. Biochem. Biophys. Res. Commun. **206:**64–76.
5. **Aldabe, R., A. Irurzun, and L. Carrasco.** 1997. Poliovirus protein 2BC increases cytosolic free calcium concentrations. J. Virol. **71:**6214–6217.
6. **Bañó-Polo, M., F. Baldin, S. Tamborero, M. A. Marti-Renom, and I. Mingarro.** 2011. N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. Protein Sci. **20:**179–186.
7. **Barco, A., and L. Carrasco.** 1995. A human virus protein, poliovirus protein 2BC, induces membrane proliferation and blocks the exocytic pathway in the yeast Saccharomyces cerevisiae. EMBO J. **14:**3349–3364.
8. **Barco, A., and L. Carrasco.** 1998. Identification of regions of poliovirus 2BC protein that are involved in cytotoxicity. J. Virol. **72:**3560–3570.
9. **Baumgärtner, A.** 1996. Insertion and hairpin formation of membrane proteins: a Monte Carlo study. Biophys. J. **71:**1248–1255.
10. **Bordier, C.** 1981. Phase separation of integral membrane proteins in Triton X-114 solution. J. Biol. Chem. **256:**1604–1607.
11. **Buchholz, U. J., S. Finke, and K. K. Conzelmann.** 1999. Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J. Virol. **73:**251–259.
12. **Claros, M. G., and G. von Heijne.** 1994. TopPred II: an improved software for membrane protein structure prediction. Comput. Appl. Biosci. (Cambridge) **10:**685–686.
13. **Cross, B. C., and S. High.** 2009. Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. J. Cell Sci. **122:**1768–1777.
14. **Cserzö, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson.** 1997. Prediction of transmembrane a-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Eng. **10:**673–676.
15. **de Jong, A. S., et al.** 2008. Functional analysis of picornavirus 2B proteins: effects on calcium homeostasis and intracellular protein trafficking. J. Virol. **82:**3782–3790.
16. **de Jong, A. S., W. J. Melchers, D. H. Glaudemans, P. H. Willems, and F. J. van Kuppeveld.** 2004. Mutational analysis of different regions in the coxsackievirus 2B protein: requirements for homo-multimerization, membrane permeabilization, subcellular localization, and virus replication. J. Biol. Chem. **279:**19924–19935.
17. **de Jong, A. S., et al.** 2002. Multimerization reactions of coxsackievirus pro-

teins 2B, 2C and 2BC: a mammalian two-hybrid analysis. J. Gen. Virol. **83:**783–793.
18. **de Jong, A. S., et al.** 2003. Determinants for membrane association and permeabilization of the coxsackievirus 2B protein and the identification of the Golgi complex as the target organelle. J. Biol. Chem. **278:**1012–1021.
19. **Engelman, D. M., and T. A. Steitz.** 1981. The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell **23:**411–422.
20. **Enquist, K., et al.** 2009. Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. J. Mol. Biol. **387:**1153–1164.
21. **García-Sáez, A. J., I. Mingarro, E. Perez-Paya, and J. Salgado.** 2004. Membrane-insertion fragments of Bcl-xL, Bax, and Bid. Biochemistry **43:**10930–10943.
22. **Gonzalez, M. E., and L. Carrasco.** 2003. Viroporins. FEBS Lett. **552:**28–34.
23. **Hedin, L. E., et al.** 2010. Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. J. Mol. Biol. **396:**221–229.
24. **Hessa, T., et al.** 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature **433:**377–381.
25. **Hessa, T., et al.** 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature **450:**1026–1030.
26. **Hirokawa, T., S. Boon-Chieng, and S. Mitaku.** 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinformatics **14:**378–379.
27. **Johnson, A. E., and M. A. van Waes.** 1999. The translocon: a dynamic gateway at the ER membrane. Annu. Rev. Cell Dev. Biol. **15:**799–842.
28. **Jones, D. T.** 2007. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics **23:**538–544.
29. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
30. **Luik, P., et al.** 2009. The 3-dimensional structure of a hepatitis C virus p7 ion channel by electron microscopy. Proc. Natl. Acad. Sci. U. S. A. **106:**12712–12716.
31. **Madan, V., A. Castello, and L. Carrasco.** 2008. Viroporins from RNA viruses induce caspase-dependent apoptosis. Cell Microbiol. **10:**437–451.
32. **Madan, V., et al.** 2007. Plasma membrane-porating domain in poliovirus 2B protein. A short peptide mimics viroporin activity. J. Mol. Biol. **374:**951–964.
33. **Martínez-Gil, L., A. E. Johnson, and I. Mingarro.** 2010. Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. J. Virol. **84:**5520–5527.
34. **Martínez-Gil, L., J. Perez-Gil, and I. Mingarro.** 2008. The surfactant peptide KL4 sequence is inserted with a transmembrane orientation into the endoplasmic reticulum membrane. Biophys. J. **95:**L36–L38.
35. **Martínez-Gil, L., et al.** 2009. Plant virus cell-to-cell movement is not dependent on the transmembrane disposition of its movement protein. J. Virol. **83:**5535–5543.
36. **Martínez-Gil, L., A. Sauri, M. A. Marti-Renom, and I. Mingarro.** 2011. Membrane protein integration into the ER. FEBS J. [Epub ahead of print.] doi:10.1111/j.1742-4658.2011.08185.x.
37. **Martínez-Gil, L., A. Sauri, M. Vilar, V. Pallas, and I. Mingarro.** 2007. Membrane insertion and topology of the p7B movement protein of melon necrotic spot virus (MNSV). Virology **367:**348–357.
38. **Monné, M., M. Hermansson, and G. von Heijne.** 1999. A turn propensity scale for transmembrane helices. J. Mol. Biol. **288:**141–145.
39. **Monné, M., I. Nilsson, A. Elofsson, and G. von Heijne.** 1999. Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. J. Mol. Biol. **293:**807–814.
40. **Nagy, A., and R. J. Turner.** 2007. The membrane integration of a naturally occurring alpha-helical hairpin. Biochem. Biophys. Res. Commun. **356:**392–397.
41. **Navarro, J. A., et al.** 2006. RNA-binding properties and membrane insertion of Melon necrotic spot virus (MNSV) double gene block movement proteins. Virology **356:**57–67.
42. **Nilsson, I., and G. von Heijne.** 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J. Biol. Chem. **268:**5798–5801.
43. **Nugent, T., and D. T. Jones.** 2009. Transmembrane protein topology prediction using support vector machines. BMC Bioinformatics **10:**159.
44. **Orzáez, M., J. Salgado, A. Gimenez-Giner, E. Perez-Paya, and I. Mingarro.** 2004. Influence of proline residues in transmembrane helix packing. J. Mol. Biol. **335:**631–640.
45. **Patargias, G., T. Barke, A. Watts, and W. B. Fischer.** 2009. Model generation of viral channel forming 2B protein bundles from polio and coxsackie viruses. Mol. Membr. Biol. **26:**309–320.
46. **Rapoport, T. A., V. Goder, S. U. Heinrich, and K. E. Matlack.** 2004. Membrane-protein integration and the role of the translocation channel. Trends Cell Biol. **14:**568–575.
47. **Rost, B., P. Fariselli, and R. Casadio.** 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci. **5:**1704–1718.
48. **Sánchez-Martínez, S., et al.** 2008. Functional and structural characterization of 2B viroporin membranolytic domains. Biochemistry **47:**10731–10739.
49. **Sandoval, I. V., and L. Carrasco.** 1997. Poliovirus infection and expression of the poliovirus protein 2B provoke the disassembly of the Golgi complex, the organelle target for the antipoliovirus drug Ro-090179. J. Virol. **71:**4679–4693.
50. **Saurí, A., P. J. McCormick, A. E. Johnson, and I. Mingarro.** 2007. Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. J. Mol. Biol. **366:**366–374.
51. **Saurí, A., S. Saksena, J. Salgado, A. E. Johnson, and I. Mingarro.** 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.
52. **Shakin-Eshleman, S. H., S. L. Spitalnik, and L. Kasturi.** 1996. The amino acid at the X position of an Asn-X-ser sequon is an important determinant of N-linked core-glycosylation efficiency. J. Biol. Chem. **271:**6363–6366.
53. **Silberstein, S., and R. Gilmore.** 1996. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. FASEB J. **10:**849–858.
54. **Skern, T., et al.** 2002. Structure and function of picornavirus proteinases, p. 199–212. *In* B. L. Semler and E. Wimmer (ed.), Molecular biology of picornaviruses. ASM Press, Washington, DC.
55. **Tamborero, S., M. Vilar, L. Martinez-Gil, A. E. Johnson, and I. Mingarro.** 2011. Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). J. Mol. Biol. **406:**571–582.
56. **van der Werf, S., J. Bradley, E. Wimmer, F. W. Studier, and J. J. Dunn.** 1986. Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. Proc. Natl. Acad. Sci. U. S. A. **83:**2330–2334.
57. **van Kuppeveld, F. J., J. M. Galama, J. Zoll, P. J. van den Hurk, and W. J. Melchers.** 1996. Coxsackie B3 virus protein 2B contains cationic amphipathic helix that is required for viral RNA replication. J. Virol. **70:**3876–3886.
58. **van Kuppeveld, F. J., et al.** 1997. Coxsackievirus protein 2B modifies endoplasmic reticulum membrane and plasma membrane permeability and facilitates virus release. EMBO J. **16:**3519–3532.
59. **van Kuppeveld, F. J., W. J. Melchers, P. H. Willems, and T. W. Gadella, Jr.** 2002. Homomultimerization of the coxsackievirus 2B protein in living cells visualized by fluorescence resonance energy transfer microscopy. J. Virol. **76:**9446–9456.
60. **Viklund, H., A. Bernsel, M. Skwark, and A. Elofsson.** 2008. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics **24:**2928–2929.
61. **Vilar, M., et al.** 2002. Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. J. Biol. Chem. **277:**23447–23452.
62. **Wimmer, E., C. U. Hellen, and X. Cao.** 1993. Genetics of poliovirus. Annu. Rev. Genet. **27:**353–436.

Reviewing Process

Overall concerns (Editor)

1- Use of BHK cells:

BHK cells have been previously used in Carrasco's lab in studies involving the expression of individual poliovirus (PV) products. As correctly stated by the Editor, these cells are not susceptible to viral infection (they lack the PV Receptor), but nonetheless posses the machinery required for synthesizing the virion components, and even to assemble infectious particles. This has been demonstrated by transfecting BHK cells with replicons that express PV mRNA (see for instance: Welnowska et al., PLoS ONE, 2011 ; 6(7):e22230). As a consequence, these cells are suitable models for investigating 2B biogenesis. This is explicitly stated now on page 14.

2- Implications for the entire polyprotein:

Our results put forward the possibility that co-translational insertion of the 2B moiety into membranes may constitute a physiologically relevant phenomenon. The fact that the inferred topology (both protein ends facing the cytosol) ensures the accessibility of viral protease cleavage sites in the P2 precursor supports such hypothesis. This is now discussed in more detail in the revised version of the manuscript (page 16)

3- Significance in the context of viral infection:

Our combined analysis predicts a marginal propensity for 2B polypeptide insertion into membranes (Table 1, Fig. 3D). On the other hand, upon viral entry, initially synthesized 2B or 2BC proteins will remain diluted into the cytosol of the infected cell. Marginal hydrophobicity together with low polypeptide concentration is predicted to reduce the probability for spontaneous insertion of 2B and 2BC into their primary target organelle: the Golgi complex (Sandoval and Carrasco L (1997) J Virol. 71:4679-93; de Jong et al. (2003) J. Biol. Chem. 278: 1012-1021). Thus, we speculate that co-translational insertion into ER membrane may be a particularly relevant phenomenon at the initial stages of the infectious cycle, during which both, the viral mRNA levels and the concentration of the translated viral proteins are low. Under those conditions the cellular protein biosynthesis and vesicular transport machineries remain functional, and the 2B and its 2BC precursor are likely synthesized as additional cell membrane proteins. From the ER, these proteins may reach the Golgi compartment membrane to fulfill regulatory and/or signaling functions that result in the disruption of Ca^{2+} homeostasis (Aldabe et al. (1997) J Virol. 71:6214-7; de Jong et al. (2008) J Virol. 82: 3782-90) and vesicular transport inhibition (Barco and Carrasco (1995) EMBO J. 17;14:3349-64; de Jong et al. (2008) J Virol. 82: 3782-90). At later stages, the canonical ER-Golgi protein-traffic pathway is no longer functional and/or required for viral replication. The massive proliferation of cell endomembranes (viroplasm) and synthesis of viral proteins results in higher effective concentrations of these components inside the infected-cell system, which may promote the spontaneous

insertion into the membrane of a significant amount of the synthesized 2B and 2BC. This possibility is discussed now on pages 17 and 18.

Specific Responses to each point raised by the Reviewers

Reviewer #1

Point 1: "In the in vivo results ... I suspect the glycosylation site was not recognized because it was too close to the transmembrane domains, however this should be explained with proper references..."

Referee's assumption is totally correct, we have included a sentence to explain this fact with the proper references in page 14.

1 (continued): "... Because all the constructs show negative results, it would have been nice to have a positive control that do show glycosylation in vivo..."

As suggested by the reviewer, we have designed a truncated version of PV 2B protein and also a construct in which the second hydrophobic region (HR2) has been deleted. Unfortunately, none of these constructs retain the ability to permeabilize cells. In addition, these truncated 2B variants are not detected by our anti-2B antibody precluding any further analysis related to the glycosylation of these small polypeptides. These results are included in the attached Figure at the end of the present letter.

Point 2: "line 29 and line 64, the sentence read as if it was expected that the protein should be co-translationally inserted in the membrane. Rephrase to "specific information regarding its biogenesis and mechanism of membrane insertion is lacking" (or mode of insertion)"

We have included the suggested rephrasing in pages 2 and 3.

Point 3: "line 276: what is meant by "isolated HR1 segments"? Do the author mean isolated HR1 segment? or isolated HR1 and HR2 segments?"

We refer in this sentence to HR1. We have corrected the wording in the current version (page 11).

Point 4: "... How about a Ncyt/Clum terminology (for luminal orientation)?

We have replaced the "*out"* abbreviation by "*lum"* throughout the manuscript, according to the referee's suggestion.

Point 5: "Assuming that a co-translational membrane insertion mechanism of 2B is indeed occuring in the context of a viral infection, I would like to see the authors discussed the implications of this model for the entire polyprotein, ..."

Regarding the P2 polyprotein the topology found in our present work would leave both the N- and C-termini of 2B facing the cytosol. This orientation is important, firstly to direct the P2 precursor to the membrane, as well as to

retain the 2BC precursor in these membranes. In addition, this topology is necessary to expose the cleavage sites recognized by PV 3C protease, in order to generate the mature products: 2A, 2B and 2C. See also previous response to the Editor.

Point 6: "Figure 2A and Figure 4A, I think it would be useful to add the theoretical size of each fusion protein, as well as the theoretical size of the fragment protected from the proteinase K."

We have included the theoretical size of the polypeptides in the corresponding figure legends. Nevertheless, it should be noted that in the case of membrane proteins, electrophoretic mobility can be significantly altered.

Reviewer #2

Related to the introductory paragraph, we would like to point out that the topology demonstrated in our work is not Ncyt/Cout as stated by the referee, but with both, the N- and C-termini facing the cytosol.

As mentioned before, this topology guarantees the proper exposure of viral protease cleavage sites within P2 precursor. In addition, co-translational insertion into the ER membrane is the likely mechanism that allows a marginally hydrophobic polypeptide (see our combined analysis), which is initially synthesized at low levels, to reach the Golgi compartment and fulfill its regulatory functions (see also previous response to the Editor).

Major concerns:

Point 1. "how could microsome system represent the ER membrane?..."

The microsomes are obtained from dog pancreas ER membranes. This system has been profusely and routinely used to perform studies on membrane protein insertion and topology, like the ones included in our work (see for instance: Hessa et al. (2005) Nature 433:377-381; Martínez-Gil et al. (2010) J. Virol. 84:5520-5527). This microsomal system provides a sensitive way of detecting the insertion or translocation of hydrophobic regions through the Sec-translocon. An obvious advantage with this technique is that the experiments are performed in the context of a biological membrane; the price one pays is that the complexity of the experimental system only allows indirect structural interpretations of the data. We have emphasized these technical advantages on page 10.

1. (continued) "...Second, the previous study has showed that the possible subcellular localization of poliovirus 2B was mainly in Golgi complex (de Jong, A.S. et al. J Virol 82, 3782-3790 (2008).) but not in ER..."

Membrane protein biosynthesis occurs at ribosomes attached to the ER membrane, subsequently membrane trafficking allows membrane proteins to

reach any further cell membrane, *i.e.* Golgi complex, plasma membrane, … Results in our paper are consistent with the use of this canonical trafficking pathway by the viroporin 2B, at least during the initial stages of infection, while it keeps functioning. We are not aware of any report on membrane protein biogenesis at the Golgi complex.

1. (continued) "…2B may not only locate on intercellular membrane system but also on cell surface… "

This comment is somehow related to the previous one, and again, we infer that during the initial stages of infection membrane traffic would handle membrane proteins fate. The fact that 2B permeabilizes cells to hygromycin B indicates that 2B is on membranes, since the inhibition of 2B to interact with membranes blocks membrane permeabilization.

Point 2. "In this study, the authors used BHKT7 cells to validate their discovery."

It is well established that PV can replicate in BHK cells. The virus cannot infect BHK cells because it is unable to attach an enter into them. However, when these cells are transfected with PV replicons, there is no inhibition of viral replication or synthesis of viral proteins (see for instance our recent publication by Welnowska et al., PLoS ONE, 2011 ; 6(7):e22230). Transfection of viral RNA into BHK cells leads also to the production of infectious virus particles. In our hands, BHK cells are excellent host to express PV proteins. In this particular case the BHK cell line employed constitutively expresses T7 RNA polymerase, giving rise to the synthesis of significant amounts of PV 2B that can be detected by western blotting upon transfection of the construct pTM1-2B. See also previous response to the Editor.

Point 3. "The authors did a good job in biochemistry. However, the authors failed to provide enough evidence to show the supposed relationship between the characters of 2B protein they discovered here and its function in poliovirus life cycle…"

Although it could be possible that the trafficking of 2B may be influenced by the expression of other viral proteins, we doubt that the topology of this protein on the membrane is modified by the expression of the rest of viral proteins. Several groups have demonstrated that the individual expression of 2B increases membrane permeability in a way akin to that observed in poliovirus infected cells. Also, the subcellular distribution of 2B is similar when individually expressed in culture cells or in virus-infected cells. Therefore, the topology of 2B would be similar in both cases.

Point 4. "…Mutants harboring deleted domain or substituted residue (as discussed by the author in line465-474) may help to illustrate the role of 2B topology in permeabilizing function."

Please, see above, response to Point 1 Referee #1 (continued).

Minor points:

Point 1. "Some information refereeing the biochemistry method needs to be further explained to the author for better understanding. "

Please see our comments above related to the Major concern 1. In addition, we have added in the Material and Methods section (page 5) a comment relative to the microsomal membranes extraction and purification process from the supplier.

Point 2. "The author is using PV1 strain which needs to be explained to the readers.".

As suggested by the referee, we have included this information on page 5.

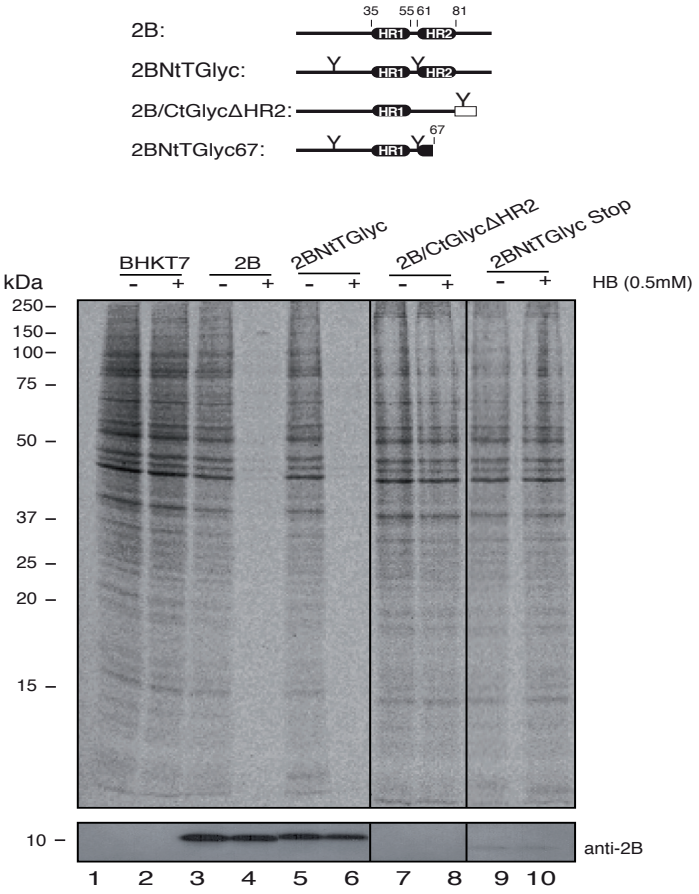
Point 3. "The HR1 and HR2 have been already predicted by previous publication (J Biol Chem 277, 40434-40441 (2002)), which should be explained in Line 67-70.

The referred work is now mentioned on page 3.

Point 4. "Several important references are missing between Line52-60.

The appropriate references have been added (page 3).

Response to Reviewer #1



Reviewing Process 2

Rebuttal

Reviewer #1:

"The picornavirus 2B protein is a well-known viroporin that associates with and modify membranes ... While some of the results are interesting and the manuscript is in general well-written, they represent a relatively small increment in our understanding of the function of this protein."

Viroporins are an important family of viral membrane proteins, but information on their membrane integration during biogenesis was lacking. We provide experimental evidence consistent with a transmembrane disposition of the 2B protein with two membrane-spanning regions and orienting both N- and C-termini on the cytosolic face of the membrane, an important conclusion that gives for the first time experimental support to the great majority of current models. In addition, because the protein inserts co-translationally into the membranes *in vitro*, we also conclude that the protein use the translocon to insert into the membranes, which is a new concept in the case of viroporins, emphasizing the role of biological membranes in the virus infection mechanism and focusing our attention on the interactions between viral and host membrane components.

"The main problem of this manuscript is that all the experiments are performed in vitro using fused proteins or truncated proteins that could affect the topology of the protein. For example, the experiments presented in Fig. 1 and 2, indicate that the association of 2b with membranes is hindered by fusion of the protein to the P2 protein. Similarly, fusion of the b2 protein to the Lep protein (in Fig. 3 and 4) inhibit the insertion a transmembrane domain (HR1) which is otherwise able to insert in the membrane on its own (Fig. 6). In fact in the case of Fig. 3 and 4, fusion of the Lep protein would force an opposite orientation of the transmembrane domain in the membranes, creating a very artificial situation. Thus the results presented in these two figures are neither informative nor convincing."

To solve this problem related with the orientation on the HR domains in the Lep model protein we have setup a new Lep-derived construct (Lep', new figure 4) in which the topology for HR1 is reversed and our current data show that in fact HR1 is properly inserted when expressed in the N_{cyt}/C_{out} orientation.

"It is only when the truncated or full-length 2B protein is used in Fig. 6 and 7 that the major conclusions can be drawn. Although the results presented in these two figures are interesting, the overall significance of the study would have been greatly improved if it was accompanied by alternative approaches to confirm the conclusions. For example, it would have been nice to see a proteinase K digestion assay using the native full-length 2B protein inserted in the microsomal membranes to confirm the proposed topology (only the transmembrane hairpin should be protected from the digestion)."

We have performed the suggested experiments but in this case the hairpin is really short (<5kDa) to be clearly detected by SDS-PAGE and, more important, the amino acid sequence in this region does not present methionine residues reducing the radioactive signal in the proteinase K experiments.

"Several possible in vivo approaches could have also been used to confirm the in vitro data"

In the current version we have included data obtained *in vivo* by transfecting BHK cells with 2B plasmids (Fig. 8). These results suggested that both *in vitro* and *in vivo* the protein acquires the same membrane-integrated structure.

Reviewer #2:

1. "The title of this MS announced that this study focus on the 2B protein of picornaviruses, however, in the text the author only mentioned 2B of polioviruses, which is only one member of the family Picornaviridae. Considering the variety of amino acid sequences, the author needs to provide more data about 2B from other viruses of this family to prove that they have the similar characters. Moreover, the author failed to provide information about the isolate of poliovirus used in this study. This will be hard to judge whether this study is representative enough even within poliovirus."

We have adapted the title to the current set of results and to the referee's suggestion. On the other hand, we have used the poliovirus Mahoney strain kindly provided by Dr. E. Wimmer (Stony Brook, USA), and this fact has been properly indicated in our current version.

2. "The author emphasized that it is through a translocon-mediated mechanism which 2B inserts into membrane (both in the title and abstract). However, only one indirect evidence was provided to prove this (Fig. 6B). Concerning that this is one of the main issue discussed in this paper, more direct experiments should be performed. For example, whether translocon induce the insertion by directly interaction? Could the author use some other methods to inhibit this machinery (siRNA or some inhibitors) to see the failure of insertion"

Data supporting the fact that viroporin 2B is inserted into the ER membrane through the translocon are provided in current figures 4, 5, 6 and 7. Nevertheless, we tuned-down our emphasis both in the title and in the abstract sections.

3. "The whole paper mainly focus on in vitro biochemistry experiments, yet none of these phenomenon have been proved in a cell model (for example, in virus-infected cells or 2B plasmid transfected cells). "

Experiments using BHK cells as a model system are included in the present version of our m/s (Fig. 8, see also response to referee #1).

Reviewer #3:

- "Can the authors shortly explain the lysate as a kind of cell free expression system? In the first two sections, 'Viroporin 2B is an integral..' and 'Insertion of the viroporin 2B...' the translocon is not mentioned, but it is assumed that it is in the lysate? Then the paragraph 'Viroporin 2B integrates...' mentions the translocon. It reads as if the translocon is now added to the experiment. This needs some clarification."

Apparently, it was not clearly expressed in our previous version the translation system used in the current experiments. These experiments were performed using rabbit reticulocyte lysates as a source of cell-free translation extracts. The translocon components are provided by the ER-derived microsomes from dog pancreas, so they are only present when microsomes are added to the translation mixtures. We believe that this issue is more clearly stressed in the current version both in the main text as well as in the Materials and Methods section.

- "Page 13, first paragraph of 'Viroporin 2B integrates...', last sentence: '... in the context of the viroporin sequence.'. This sentence is hard to understand.

We have modified this part of the m/s to clarify the protein topology results.

- "At some stage the phrase 'TM segment' is used then 'HR1' and 'HR2'. This needs to be made consistent.

We have used the term 'TM segment' to refer to sequences that have been experimentally validated to span the membrane, while 'HR1 and HR2' refer to the 2B sequence regions that include residues 35-55 and 61-81, respectivelly.

- "It would help to see somewhere the entire amino acid sequence of 2B".

Viroporin 2B amino acid sequence is now included in Fig. 1.

- "In the text the description of the figures starts mostly with the part B of the figures rather than with A. For Figure 7 the first part being referred to in the text is part C. This situation needs to be cleared."

We have corrected this situation in the current version.

Chapter 3:

Charge Pair Interactions in Transmembrane Helices and Turn Propensity of the Connecting Sequence Promote Helical Hairpin Insertion

jmb Article

Article 3

Charge Pair Interactions in Transmembrane Helices and Turn Propensity of the Connecting Sequence Promote Helical Hairpin Insertion

Manuel Bañó-Polo[1], Luis Martínez-Gil[1], Björn Wallner[2], José L. Nieva[3], Arne Elofsson[4] and Ismael Mingarro[1]

*1 - **Departament de Bioquímica i Biologia Molecular,** Universitat de València, E-46100 Burjassot, Spain*
*2 - **Department of Physics,** Chemistry and Biology, Swedish e-Science Research Center, Linköping University, SE-581 83 Linköping, Sweden*
*3 - **Unidad de Biofísica (CSIC-UPV/EHU) and Departamento de Bioquímica,** Universidad del País Vasco (UPV/EHU), E-48080 Bilbao, Spain*
*4 - **Department of Biochemistry and Biophysics,** Science for Life Laboratory, Stockholm Bioinformatics Center, Center for Biomembrane Research, Swedish e-Science Research Center, Stockholm University, SE-106 91 Stockholm, Sweden*

Correspondence to Ismael Mingarro: *Ismael.Mingarro@uv.es*
http://dx.doi.org/10.1016/j.jmb.2012.12.001
Edited by J. Bowie

Abstract

α-Helical hairpins, consisting of a pair of closely spaced transmembrane (TM) helices that are connected by a short interfacial turn, are the simplest structural motifs found in multi-spanning membrane proteins. In naturally occurring hairpins, the presence of polar residues is common and predicted to complicate membrane insertion. We postulate that the pre-packing process offsets any energetic cost of allocating polar and charged residues within the hydrophobic environment of biological membranes. Consistent with this idea, we provide here experimental evidence demonstrating that helical hairpin insertion into biological membranes can be driven by electrostatic interactions between closely separated, poorly hydrophobic sequences. Additionally, we observe that the integral hairpin can be stabilized by a short loop heavily populated by turn-promoting residues. We conclude that the combined effect of TM–TM electrostatic interactions and tight turns plays an important role in generating the functional architecture of membrane proteins and propose that helical hairpin motifs can be acquired within the context of the Sec61 translocon at the early stages of membrane protein biosynthesis. Taken together, these data further underline the potential complexities involved in accurately predicting TM domains from primary structures.

Introduction

Multi-spanning membrane proteins (those including two or more membrane-spanning segments) are important for many biological functions. The basic structural unit of such membrane proteins is a hydrophobic α-helix. In folded proteins, these individual helix-forming sequences are engaged in a rich network of interactions with other helices. Whereas individual helices are formed in response to main-chain hydrogen bonding and the hydrophobic effect, other interactions must be responsible for side-to-side assembly. Such interactions might include hydrophobic packing, electrostatic effects, turns between helices, and binding to components placed in the aqueous environments that surround the membrane.

α-Helical hairpins, consisting of a pair of closely spaced transmembrane (TM) helices that are connected by a short extramembrane or interfacial turn, are the simplest structural motifs found in multi-spanning membrane proteins.[1] This motif is thought to occur relatively frequently in integral membrane proteins and may serve as an important structural and/or functional element.[2]

The insertion of most helical plasma membrane proteins occurs co-translationally, whereby protein

0022-2836/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.

synthesis and integration into the membrane are coupled. For the integration of individual TM sequences into the membrane, it is expected that TM segments will preadopt a helical state,[3,4] due to the significant free-energy penalty of embedding an exposed polypeptide backbone into the hydrophobic membrane core.[5] Similarly, the formation of interhelical hydrogen bonds facilitates the integration of polar residues present in adjacent TM regions. Then, there is no doubt that hydrogen bond interactions can play key roles in helical hairpin stabilization.

While many studies addressing the formation of helical hairpins in membranes have been carried out on model hydrophobic TM segments,[6–9] naturally occurring helical hairpins are not always highly hydrophobic,[10] and the role of helix–helix interactions and turn propensities of the residues interconnecting the two helices in their folding and stability is poorly understood. We previously showed that poliovirus (PV) 2B, which is a small protein involved in PV virulence, is a double-spanning integral membrane protein, in which the two TM segments are interconnected by a short turn forming a putative helical hairpin.[11] As a first step towards understanding its biogenesis, we demonstrated that *in vitro* PV 2B integrates into the endoplasmic reticulum (ER) membrane through the translocon.[11]

Here, we present a detailed investigation on structural determinants underlying helical hairpin formation in the viral membrane protein 2B. Using an *in vivo*-like translation–glycosylation system of the naturally occurring helical hairpin from PV 2B, we have determined the importance of helix–helix interactions for hairpin formation. In addition, we show that the hairpin structure is stabilized by the turn propensity of the amino acid residues in the short loop between the two TM helices. Our results suggest that integral helical hairpins may form in biological membranes driven by electrostatic interactions between marginally hydrophobic sequences and be additionally stabilized by short, tight connecting turns.

Results

Insertion of the viroporin 2B hairpin region into biological membranes

We have recently shown that *in vitro* PV 2B product inserts into the ER membrane as a double-spanning integral membrane protein with an N-/C-terminal cytoplasmic orientation.[11] Such topology is attained upon insertion of a helical hairpin whose constituent TM helices are marginally hydrophobic (Fig. 1a). *In vitro* synthesis of several truncated protein versions indeed put forward that the two hydrophobic regions cooperate to insert into the ER-derived microsomal membranes. Here, we explore the structural grounds for such effect.

As in our previous study of membrane insertion of the PV 2B,[11] we used a well-characterized *in vitro* experimental system based on glycosylation[12] that accurately reports the integration of TM regions into microsomal membranes. Upon insertion, the oligosaccharyl transferase (OST) enzyme modifies the protein of interest. OST adds sugar molecules to an NX(S/T) consensus sequence,[14] with X being any amino acid except proline,[15] after the protein emerges from the translocon channel. Glycosylation of a protein region synthesized *in vitro* in the presence of microsomal membranes therefore indicates the exposure of this region to the OST active site on the luminal side of the ER membrane. When assayed independently, the two hydrophobic regions of the PV 2B did not span the ER-derived membranes,[11] as expected according to the predicted apparent free energy of insertion (Fig. 1a). It has been shown previously that, in some cases, a neighboring TM helix can promote membrane insertion of a marginally hydrophobic TM region[16–19] and that there is a correlation between the polarity of a TM helix and its interaction area with the rest of the protein.[20] Therefore, we investigated the insertion of the full α-helical hairpin region (residues 35–81) in this *in vitro* translation system.

In our first experimental setup, the helical hairpin region was introduced into the "host" protein leader peptidase (Lep) (Fig. 1b), which contains two TM helices (H1 and H2) and a large lumenally exposed C-terminal domain (P2). In this first Lep construct, the 2B hairpin sequence (residues 35–81, Fig. 1a) was placed near the middle of the P2 domain (Fig. 1b) and was flanked by two engineered NXT acceptor sites for *N*-linked glycosylation (G1 and G2), with the G2 site located immediately downstream the hairpin region (see Materials and Methods). It has previously been demonstrated that efficient glycosylation occurs when the acceptor Asn is ~12–14 residues away from the membrane.[21,22] If the hairpin is translocated across the membrane, both G1 and G2 sites will be modified by the lumenally oriented OST; if the helical hairpin is inserted into the membrane, only G1 will receive a glycan (Fig. 1b). If one of the two hydrophobic regions is inserted, only G1 will be modified, but in that case, the large P2 domain will be non-translocated across the microsomal membrane. In this way, a single glycosylation suggests either correct hairpin integration (Fig. 1b, right) or the integration of only one hydrophobic region, whereas double glycosylation reports the non-integration capability of the hairpin region (Fig. 1b, left). Single glycosylation of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep

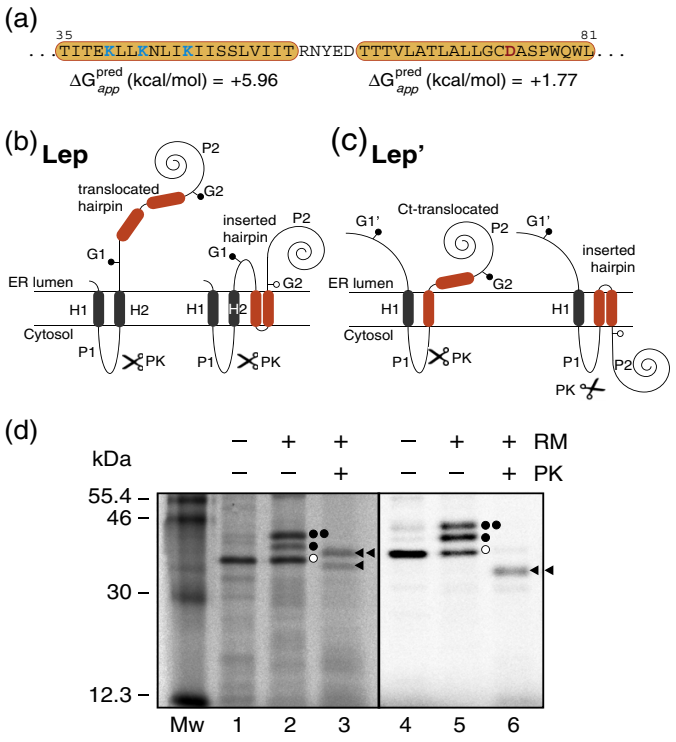


Fig. 1. In-block insertion of PV 2B α-helical hairpin (residues 35–81) into microsomal membranes using a model protein. (a) Sequences for the predicted TM helices (boxes) and connecting turn. The predicted free energy of insertion (Δ*G*) apparent values (estimated using the Δ*G* prediction server at http://dgpred.cbr.su.se[12,13]) is indicated at the bottom of each TM segment. In this algorithm, positive values are indicative of translocation across the membrane (i.e., absence of stable insertion). Lysine and aspartic acid residues are shown in blue and red, respectively. Schematic representations of the Lep (b) and Lep′ (c) constructs used to report insertion into the ER membrane of 2B helical hairpin. In the Lep construct (b), the helical hairpin under study is inserted into the P2 domain and flanked by two artificial glycosylation acceptor sites (G1 and G2). Recognition of the hairpin by the translocon machinery as an integrating domain locates G1 and G2 in the luminal side of the ER membrane, but the short distance to the membrane prevents G2 glycosylation. The Lep chimera will be doubly glycosylated when the hairpin is translocated into the lumen of the microsomes, and insertion of only one hydrophobic region would render singly glycosylated proteinase K (PK)-digestible molecules. (c) In this Lep′ construct, PV 2B hairpin replaces the H2 domain from Lep. The glycosylation acceptor site (G2) will be modified only if translocated across the membrane, while the G1′ site, embedded in an extended N-terminal sequence, is always glycosylated. (d) *In vitro* translation in the presence (+) or absence (–) of RMs and PK. Bands of non-glycosylated protein are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. The protected singly and doubly glycosylated H2/helical hairpin/P2 fragments are indicated by one and two black triangles, respectively. Lane Mw contains radioactive molecular mass markers as indicated.

expressed in the absence of microsomes, and around 5 kDa in the case of double glycosylation.

As shown in Fig. 1d, *in vitro* synthesis of this construct yielded single and double glycosylated forms of the protein (lanes 1 and 2). Proteinase K treatment of this translation reaction mixtures produced two main bands corresponding to P1-digested molecules (Fig. 1d, lane 3), which contained the second TM segment from Lep (H2) plus the hairpin of 2B and the P2 domain either singly glycosylated (lower band) or doubly glycosylated (upper band). It should be noted that the G2 acceptor site is located too close to the C-terminal region of the α-helical hairpin (see Materials and Methods) to be efficiently glycosylated (Fig. 1b), since, as previously demonstrated, the acceptor site should be located away from the membrane interface for efficient glycosylation.[21,22] To facilitate the interpretation of these experimental results, it is important to mention that short interconnecting turns like the one present in the 2B α-helical hairpin are poor substrates for proteinase K digestion.[23] Then, the presence of a significant proportion of singly glycosylated molecules is indicative of a hairpin insertion.

Next, we tested the 2B hairpin arranged according to its predicted topology[11] using a different setup (Fig. 1c). In the Lep-derived construct (Lep′) used to measure hairpin insertion propensity in this case, the 2B hairpin sequence replaced the Lep H2 domain (Fig. 1c). The glycosylation site (G2) located in the beginning of the P2 domain will be modified only if this C-terminus (Ct) domain is translocated across the membrane, while an engineered G1′ site, embedded in an extended N-terminus sequence of 24 amino acids, is always glycosylated. As previously observed, when assayed in the Lep′ construct, the first hydrophobic region inserted efficiently into the membrane (up to 60% of the molecules were

doubly glycosylated; see Fig. 4 in Ref. 11). However, synthesizing the helical hairpin in this Lep′ construct in the presence of microsomal membranes, the doubly glycosylated band diminished to 34% of the molecules (Fig. 1d, lane 5), indicating a significant achievement of hairpin insertion reported as singly glycosylated molecules (see Fig. 1c for a scheme). Furthermore, proteinase K digestions of these samples (Fig. 1d, lanes 4–6) render protected forms derived from the doubly glycosylated molecules. Overall, these data suggest that, in the context of Lep-derived model proteins, the two hydrophobic regions of the viroporin 2B can insert into the membrane when expressed as an α-helical hairpin. They further emphasize that the preferred orientation is with the N-/C-termini facing towards the cytosolic side (Fig. 1c).

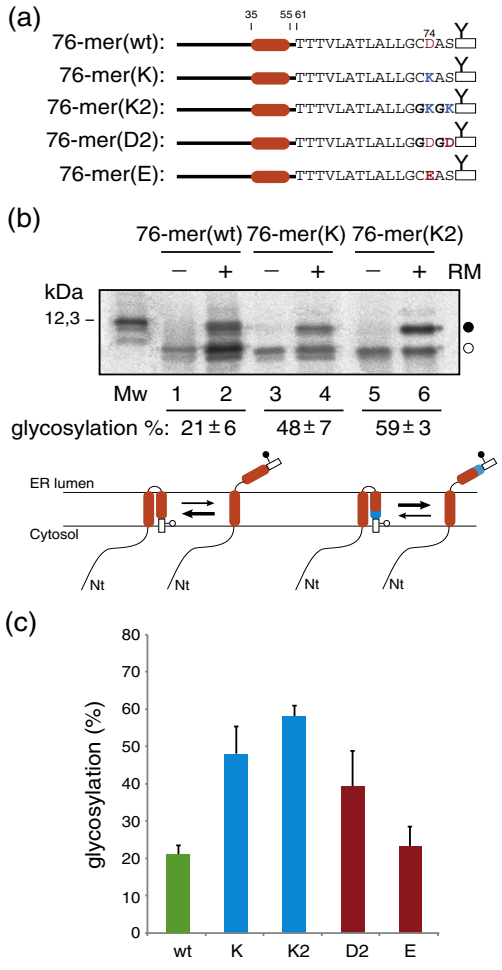


Fig. 2. Insertion of PV 2B-derived truncated proteins. (a) Structural organization of the 76-mer truncated constructs. The N-glycosylation site added as a C-terminal reporter tag (rectangles) is highlighted by a Y-shaped symbol. Mutant residues are shown in boldface and acidic and basic residues are highlighted in red and blue, respectively. (b) *In vitro* translations were performed in the presence (+) and in the absence (–) of RMs as indicated. Non-glycosylated and singly glycosylated proteins are indicated by a white and black dot, respectively. A scheme for the equilibrium between non-glycosylated and glycosylated forms for the wild type (left) and the lysine-containing constructs (right) is shown at the bottom, where the presence of lysine residues at the Ct is highlighted in blue and the thickness of the arrowhead lines refers to the prevalence of each form. (c) *In vitro* glycosylation of 76-mer truncated proteins. The level of glycosylation is quantified from SDS-PAGE gels by measuring the fraction of glycosylated (f_g) *versus* glycosylated plus non-glycosylated (f_{ng}) molecules, $p=(f_g)/(f_g)+(f_{ng})$. Data correspond to averages of at least three independent experiments and error bars show standard deviations. Lane Mw contains radioactive molecular mass markers as indicated.

Effects on helical hairpin formation of a single Asp residue in the second hydrophobic region

A phylogenetic analysis of picornavirus 2B proteins has highlighted the presence of cationic residues in the first hydrophobic α-helix and, in the case of several genera, the existence of an aspartic acid residue within the second hydrophobic segment.[24]

In our attempts to identify possible helix–helix interactions that stabilize hairpin formation, we focused first on the aspartic residue mentioned above since this appeared to be a likely candidate to promote electrostatic interactions between the TM helices. Because *N*-glycosylation acceptor sites are absent from the PV 2B sequence, we added a C-terminal glycosylation tag that has been proven to be efficiently modified[25] and synthesized the first 76 residues of the PV 2B (76-mer), which contained the native aspartic residue (Asp74) (Fig. 2a; see also Fig. 6 in Ref. 11), to emphasize the role of this polar residue. In these truncated polypeptides, glycosylation scores translocation of the Ct region across the ER membrane, while non-glycosylated protein bands correspond to insertion of this Ct region (see schemes in Fig. 2b).

To prove the role of Asp74 within these truncated protein forms, we replaced this residue by a lysine residue [76-mer(K) and 76-mer(K2)]. As shown in Fig. 2b, the glycosylation level for these mutants increased significantly compared to the wild-type truncated sequence (lanes 1 and 2), suggesting that the presence of the lysine residues complicates membrane insertion of the second hydrophobic region (see equilibrium schemes at the bottom of Fig. 2b). A similar effect was found when a second aspartic acid residue was included [76-mer(D2)]

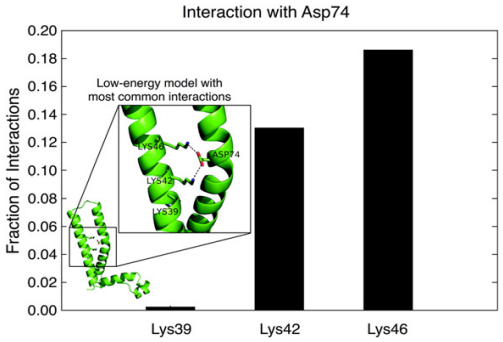


Fig. 3. Observed interactions of Asp74 among low-energy models generated by the Rosetta membrane protocol (see Materials and Methods). Clearly, it is energetically favorable for the hairpin to hide the polar groups of the Asp and Lys residues. In particular, the polar group of Asp74 can be shielded from the hydrophobic lipid environment by interactions with one of the three Lys in the first helix. The data analysis shows that interactions between Asp74 and either Lys42 or Lys46 are found much more frequently than interactions between Asp74 and Lys39, which are almost never observed. The inset shows one of the models where interactions are seen between Asp74 and both Lys42 and Lys46.

(Fig. 2c). The most likely explanation for the disruptive effect of this second aspartic acid residue (position 76) is that it lies on the side of the helix that faces the lipid (two positions away from Asp74 or ~200°), imposing an unsurpassable penalty for membrane insertion of the second helix. Interestingly, replacement of Asp74 by glutamic residue rendered a glycosylation level similar to the native aspartic residue (Fig. 2c).

Position-specific effects on helical hairpin formation by Lys–Asp pairs

We next focused on the positively charged residues present in the first hydrophobic region of the PV 2B sequence (residues 35–55, Fig. 1a). The protein sequence in this region contains three lysine residues Lys39, Lys42, and Lys46. To find out which of these residues would be involved in electrostatic interactions with Asp74, we generated possible structural models using the Rosetta membrane protocol.[26] These models clearly predicted a preference for Lys46 and Lys42 to interact with Asp74. Among the five largest structural clusters obtained, as compared to Lys39, interactions of Lys42 and Lys46 with Asp74 are 53 and 75 times more frequent, respectively (Fig. 3).

In a series of experiments performed in hairpin truncations (76-residue-long polypeptides, 76-mers), we separately replaced each lysine residue by an aspartic acid residue. As expected, K39D mutation exhibited a glycosylation level indistinguishable from the wild-type sequence (data not shown), indicating that it does not perturb hairpin insertion and confirming our computational predictions. On the contrary, as shown in Fig. 4a, replacement of Lys42 by a negatively charged aspartic acid residue (K42D mutant) substantially increased glycosylation (compare lanes 1 and 2 with lanes 5 and 6), indicating that Lys42 is involved in some electrostatic interactions between the two adjacent helices. In fact, these interactions can be partially restored as reported with the double mutant K42D/D74K, which is only weakly glycosylated (Fig. 4a, lanes 3 and 4). Thus, hairpin capacity for inserting into the ER membrane was almost unaffected by swapping charges between helices, which underscores the involvement of these residues in inter-helical electrostatic interactions.

The same set of constructs was used to test Lys46 involvement in these electrostatic interactions (Fig. 4b). In this set of experiments, a similar effect was observed: the single mutant K46D was efficiently glycosylated (C-terminal tag translocated across the membrane, lanes 1 and 2), whereas double mutant K46D/D74K was poorly glycosylated (lanes 3 and 4), indicating again that the interactions between the two hydrophobic regions were restored in this later construct.

To confirm that these interactions are relevant in the context of the 2B full-length protein, we separately mutated both Lys42 and Lys46 to aspartic acid and experimentally determined their membrane disposition. The synthesis of Ct-tagged 2B full-length protein (FL) in the presence of rough microsomal membranes (RMs) resulted in glycosylation in $<5\%$ of the molecules (Fig. 5, lanes 5 and 6). Both K42D and K46D mutants consistently showed a significantly increased glycosylation level, suggesting that the more efficient hairpin integration observed in the native 2B protein might be promoted by electrostatic interactions between Lys–Asp pairs placed in the adjacent TM segments.

Influence of turn residues on hairpin stabilization

Next, we focused on the turn between the two TM segments as an additional source of stability of the helical hairpin motif. To investigate the role of the residues interconnecting both 2B TM segments, we performed an alanine-scanning mutagenesis approach with the full-length Ct-tagged 2B protein. Alanine was used because it carries a non-bulky and chemically inert side chain and, more importantly, because it has a very low turn-promoting effect.[27] Thus, we individually replaced residues 56 to 60 with alanine. The results of these experiments are shown in Fig. 6a. Interestingly, a gradual effect was observed. Hence, Y58A and D60A mutants displayed glycosylation levels similar to that of native full-length PV 2B (compare lanes 4 and 6 in Fig. 6a

with lane 6 in Fig. 5), suggesting that individual replacement of Tyr58 or Asp60 by the poor turn-promoting alanine residue does not preclude hairpin integration. However, Glu59 and especially Arg56 and Asn57 single amino acid replacements rendered a clear increase in the glycosylation efficiency of these constructs.

Furthermore, we collected all turns in helical hairpins from predicted integral membrane proteins of viral origin (see Materials and Methods). A total of 3654 such turns were found. Turn length and amino acid frequencies were calculated for these. In Fig. 7, we show that five-residue turn helical hairpins, like the one observed in the PV 2B sequence, are among the most frequent in viral membrane, only surpassed by turns of length two.

Finally, we sought to test if there is any combined effect on the helical hairpin integration between the charge pairs found in the hydrophobic regions and the turn residues. To this end, we designed double mutations including the more sensitive residue Asn57 (N57A), which appears to play an important role in hairpin formation, and the lysines and aspartic acid residues involved in the electrostatic interactions found above. As shown in Fig. 6b, the glycosylation efficiencies for these double mutants were indeed higher compared to the single mutants analyzed above (Figs. 5 and 6a) and demonstrated an additive effect of the different protein domains in terms of helical hairpin stabilization.

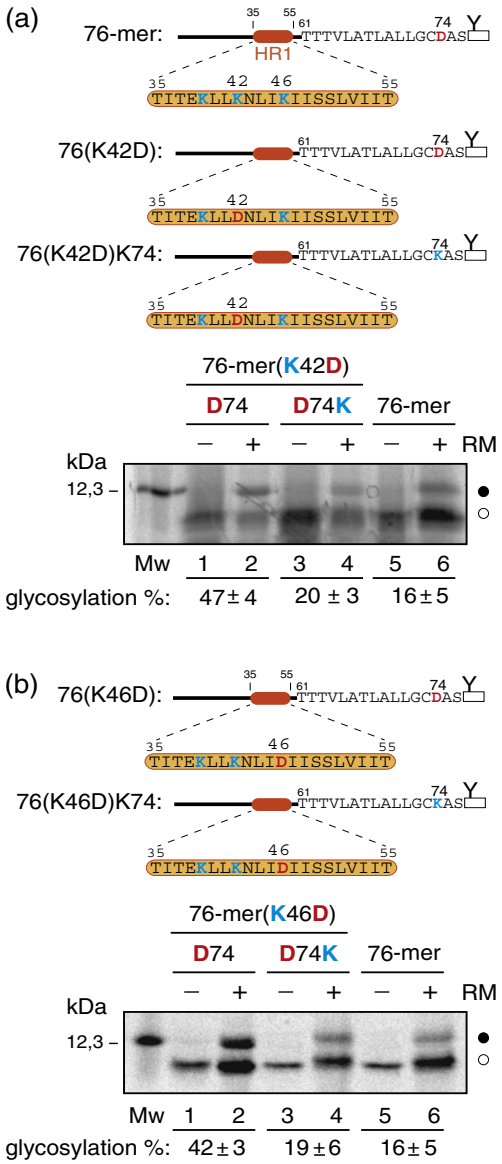


Fig. 4. Effects of Lys42–Asp74 (a) and Lys46–Asp74 (b) charge pairs on the integration of 2B truncated molecules into the ER membrane. Schematic representations of the Ct-tagged 2B 76-mer truncated constructs are shown on top of each panel. Bottom: *in vitro* translation in the presence (+) or absence (–) of RMs of 76-residue truncated molecules carrying wild-type (Asp74, lanes 1, 2, 5, and 6) or mutant D74K (lanes 3 and 4) sequences. Glycosylation levels are quantified as in Fig. 2, and non-glycosylated and singly glycosylated proteins are indicated by a white and black dot, respectively. Data correspond to averages of at least three independent experiments. Lane Mw contains radioactive molecular mass markers as indicated.

Discussion

Helical hairpins appear to be extremely common in multi-spanning integral membrane proteins and is thus an essential structural motif to our understanding of membrane protein folding and topology. In these structural motifs, the insertion of a polypeptide segment containing polar amino acids can be facilitated by the insertion of a closely spaced, more hydrophobic region.[29] Then, the membrane-buried polar groups are saturated with internal hydrogen bonds and salt bridges, and nonpolar side chains are preferentially exposed on the bilayer-facing surfaces, balancing the physical constraints imposed by the nonpolar core of the membrane bilayer and resulting in thermodynamic stability of the multi-spanning membrane protein.

In the hydrophobic environment of the membrane core, van der Waals helix–helix packing, hydrogen bonding, and ionic interactions are the governing contributors to multi-spanning membrane protein assembly. These interactions can be modulated by the sequence context and the lipid bilayer

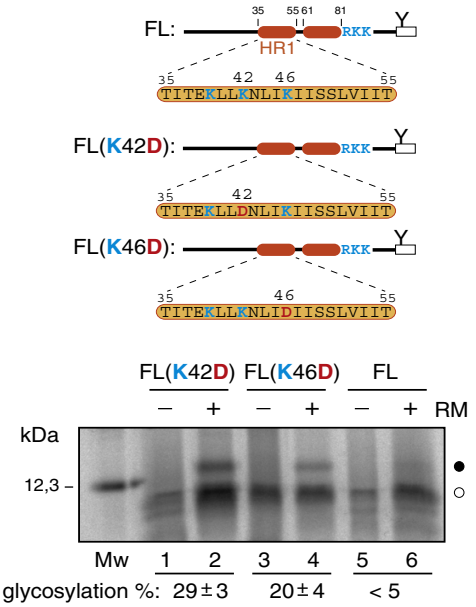


Fig. 5. Effects of mutations in the first hydrophobic segment on the integration of full-length 2B into the ER membrane. Structural organizations of full-length PV 2B constructs are shown on top, with the flanking basic residues Arg82, Lys83, and Lys84 highlighted in blue. Bottom: *in vitro* translation in the presence (+) or absence (–) of RMs of full-length molecules carrying wild-type (lanes 5 and 6) or mutant K42D (lanes 1 and 2) and K46D (lanes 3 and 4) sequences. The level of glycosylation is quantified as in Fig. 2, and non-glycosylated and singly glycosylated proteins are indicated by a white and black dot, respectively. Data correspond to averages of at least three independent experiments. Lane Mw contains radioactive molecular mass markers as indicated.

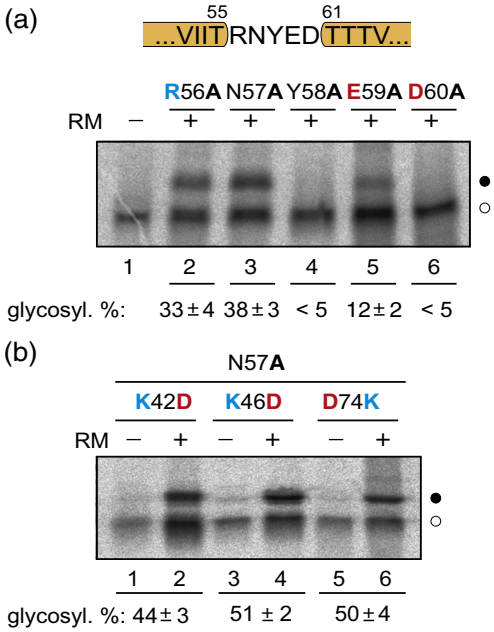


Fig. 6. Effects of mutations in the turn region. (a) Ala-scanning mutagenesis of the residues located in the turn region. *In vitro* translations were performed in the presence (+) or in the absence (–) of RMs as indicated. Turn region amino acid sequence is shown on top. (b) Additive effect of mutations in the turn and in the charged residues involved in pairwise interactions. *In vitro* translations were performed in the presence (+) or in the absence (–) of RMs as indicated. Data correspond to averages of at least three independent experiments.

properties.[30] Although hydrophobic interactions between TM segments are more abundant, hydrogen bonding or salt-bridge formation between membrane-spanning charged residues is essential to drive membrane protein folding while, at the same time, reducing the unfavorable energetics of inserting charged residues into the membrane.[31] Hence, interactions between polar residues in adjacent TM segments have been shown to favor membrane insertion.[32–34] PV 2B comprises two TM segments harboring conserved charged residues.[24] Individual integration of these sequences in a model protein construct (Lep) proved to be inefficient.[11] However, the glycosylation pattern found in this system when both regions were expressed in-block (including their native turn region) suggests that the helical hairpin components cooperate to facilitate their insertion into the membrane (Fig. 1). We show here that a central component for the stabilization of the helical hairpin is the charged residues found in the TM regions. Structure predictions placed lysine residues 42 and 46 in the first TM segment and Asp74 in the second TM segment at optimum positions to stabilize the helical hairpin conformation by inter-helical hydrogen bonding (Fig. 3). Experimentally, we found that the negatively charged residue (Asp74) appears to be relevant for hairpin integration (Fig. 2). Furthermore, Lys42 and Lys46 have an impact on hairpin integration of both truncated (Fig. 4) and full-length proteins (Fig. 5). In all these cases, substitutions of the lysine residues by aspartic acid residues have a significant impact on the glycosylation level of wild-type downstream sequences, whereas similar mutations on a background where the Asp74 was replaced by a lysine residue restored the glycosylation levels of native sequences. It is important to mention that all these mutations have a higher impact on the glycosylation level of the truncated molecules than on the full-length constructs. Interestingly, it has been

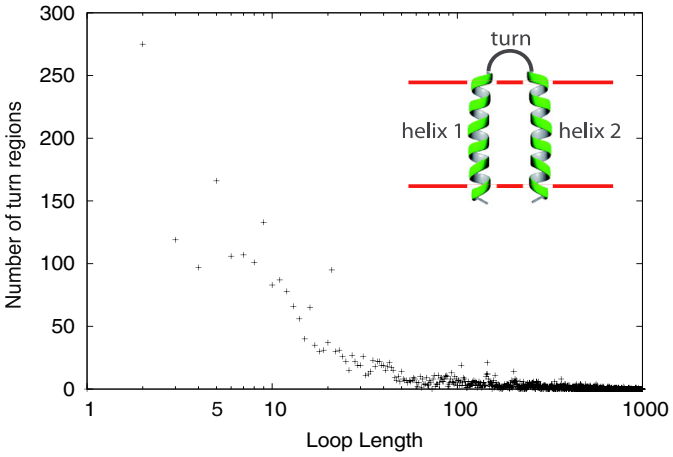


Fig. 7. Distribution of the turn lengths among viral membrane proteins. Inset: schematic diagram of a helical hairpin. Turn residues are defined by using the topcons-single prediction algorithm.[28]

previously demonstrated that extramembranous positively charged residues placed near the cytosolic end of TM segments in membrane proteins promote membrane insertion of precedent hydrophobic helices.[19,35] Inspection of the PV 2B sequence reveals the presence of three positively charged flanking residues adjacent to the helical hairpin C-terminal end (see Fig. 5, top), which could then contribute to the higher integration efficiency for the full-length protein compared to truncated molecules where these flanking positively charged residues are absent.

One of the aims of our study was to search for patterns to aid the prediction of helical hairpins from viral membrane protein sequences. The two-residue turns are most common in our viral membrane protein collection, but five-residue turns (such as the one found in PV 2B) are the second most abundant (Fig. 7).

The first two residues in the turn (Arg56 and Asn57) showed to be more sensitive to alanine replacement (Fig. 6a). Interestingly, asparagine and arginine residues together with prolines have been found to behave as the strongest turn-promoting residues when placed in the middle of a 40-residue polyleucine segment.[27,36] In addition, multiple sequence analysis of picornavirus 2B protein sequences highlighted Arg56 and Asn57 as fully conserved residues,[37] suggesting a relevant role in protein function. We also find that the contribution of turn residues can be combined with the interaction between charged residues appropriately located to stabilize the helical hairpin conformation by inter-helical hydrogen bonding, since the higher levels of glycosylation have been obtained when perturbed both, turn and charged residues at the same time (Fig. 6b). Similarly, structural perturbations introduced by extracellular loop mutations in a cystic fibrosis TM conductance regulator appeared to affect hairpin conformation to a lesser extent in the context of more stable TM helix–helix interactions,[38] highlighting the interplay between helix–helix interactions and turn contribution in modulating hairpin formation.

Our observation that Lys–Asp pairs together with turn-promoting residues stabilize a TM helical hairpin conformation supports helical hairpin formation at the initial stages of membrane protein biogenesis, that is, within the translocon. In this regard, two TM segments have been found to accumulate within or adjacent to the translocon before final integration into the lipid bilayer.[39–43] Furthermore, molecular dynamics simulations of the translocon in a fully solvated lipid bilayer showed that two helices can coexist within the translocon (J. Gumbart, personal communication). An interesting possibility arising from these observations is that the translocon may facilitate membrane protein integration by allowing efficient polar interactions between consecutive TM segments harboring charged residues and thereby stabilizing them in the nascent polypeptide. In this particular case, the lysine residues located in the first TM segment may participate in keeping this domain in the translocon until the second TM segment reaches this location. If this were the case, helical hairpin formation inside the translocon might facilitate partitioning into the lipid bilayer by shielding the charged amino acids that would otherwise constrain membrane integration efficiency. Although further studies are needed to unravel the details of this hypothesis, our results suggest that very specific helix–helix interactions can be formed within the context of the translocon and that such interactions together with turn constraints can have a dramatic effect into the poorly understood mechanism of multi-spanning membrane protein folding.

Materials and Methods

Enzymes and chemicals

All enzymes as well as plasmid pGEM1 and rabbit reticulocyte lysate were from Promega (Madison, WI). SP6 RNA polymerase and ER rough microsomes from dog pancreas were from tRNA Probes (College Station, TX). [^{35}S]Met and ^{14}C-methylated markers were from GE Healthcare. Restriction enzymes and Endoglycosidase H were from Roche Molecular Biochemicals. The DNA plasmid, RNA cleanup, and PCR purification kits were from Qiagen (Hilden, Germany). The PCR mutagenesis kit QuikChange was from Stratagene (La Jolla, CA). All the oligonucleotides were from Thermo (Ulm, Germany).

DNA manipulations

The helical hairpin from PV 2B was introduced into the modified Lep sequence from the pGEM1 plasmid[12,44] between the SpeI and KpnI sites by PCR-amplified PV 2B sequence containing residues 32–81 with primers containing appropriate restriction sites. After PCR amplification, the PCR products were purified, digested, and ligated to the corresponding Lep vector digested with the same enzymes. In this construct, the acceptor asparagine is located seven residues downstream the end of the 2B hairpin sequence (…WL^{81}QVPGQQNAT…, where L^{81} is the Ct residue of the hairpin and the glycosylation site is underlined). The Lep′ construct carried one glycosylation acceptor site in positions 3–5 of an extended sequence of 24 residues previously described.[45] The same strategy was used to insert PV 2B hairpin sequence by PCR into the Lep′ construct.

Full-length 2B sequence (without a stop codon) fused to the P2 domain of the *Escherichia coli* Lep in a pGEM1 plasmid was performed as described previously.[11,46] The K42D, K46D, and D74K site-directed mutagenesis were done using the QuikChange mutagenesis kit from Stratagene following the manufacturer's protocol. All DNA manipulations were confirmed by sequencing of plasmid DNAs. For the 76-residue-long truncated molecules (76-mers), replacement of Asp74 by lysine, alanine, glycine, and glutamic acid as well as mutants 76-mer(K2) and 76-mer(D2) was performed by using reverse primers with the appropriated sequences in the PCR amplifications (see below).

Expression *in vitro*

Lep construct with 2B helical hairpin was transcribed and translated as previously reported.[44,47]

Full-length 2B DNA was amplified from 2B/P2 plasmid using a reverse primer with an optimized Ct glycosylation tag[25] and a stop codon at the end of the 2B sequence (2B-derived expressions). Truncated 2B DNA was amplified by PCR from 2B/P2 plasmid using reverse primers with an optimized Ct glycosylation tag and a stop codon at the end annealing at specific position to obtain the desired polypeptide length (76 residues). The transcription of the DNA derived from pGEM1 plasmid was done as previously described.[48] Briefly, the transcription mixture was incubated at 37 °C for 2 h. The mRNAs were purified using a Qiagen RNeasy cleanup kit and verified on a 1% agarose gel.

In vitro translation of *in vitro* transcripted mRNA was done in the presence of reticulocyte lysate, [^{35}S]Met, and dog pancreas microsomes as described previously.[48,49] After polypeptide synthesis membranes were collected by ultracentrifugation and analyzed by SDS-PAGE, gels were finally visualized on a Fuji FLA3000 phosphorimager using the Image Gauge software.

For the proteinase K protection assay, the translation mixture was supplemented with 1 μL of 50 mM $CaCl_2$ and 1 μL of proteinase K (4 mg/mL) and then digested for 20 min on ice. Adding 1 mM PMSF before SDS-PAGE analysis stopped the reaction.

Computational modeling

Rosetta membrane[26] was used to generate 10,181 conformations of the helix hairpin from structural fragments and the topology prediction by Topcons (topcons ref: PMID: 19429891). All conformations were clustered on the helix–helix region using a 2-Å cutoff. The models belonging to the five largest clusters (20% of all clustered models) were subjected to all-atom refinement using the *relax* application in the Rosetta package. The all-atom conformational space around each model was explored by performing 250 independent all-atom refinement runs for each model before the interactions between Asp74 and Lys39, Lys42, and Lys46 were analyzed. Asp and Lys were considered to interact if the polar groups of their side chains were within 3.5 Å.

Sequence analysis of viral TM proteins

Potential viral proteins containing membrane hairpins were detected in the following way. All viral sequences were downloaded from the refseq.[50] The topcons-single algorithm (ref: PMID 21493661) was then used to predict the topology of all 81,932 viral proteins. Of these proteins, 13,848 were predicted to be TM and 3654 of these had two predicted TM helices. These proteins were used to analyze the length and composition of the loops.

Acknowledgements

This work was supported by grants BFU2009-08401 (to I.M.) and BIO2011-29792 (to J.L.N.) from the Spanish Ministry of Science and Innovation (MICINN, European Regional Development Fund supported by the EU), ACOMP/2012/226 from the Generalitat Valenciana (to I.M.), and GIU06/42 from the Basque Government (to J.L.N.). A.E. was funded by the Swedish National Research Center, Vinnova, and EU FP7 through the EDICT grant (FP7-HEALTH-F4-2007-201924). M.B.-P. was the recipient of an Formación de Profesorado Universitario fellowship from the Spanish Ministry of Education. B.W. was funded by the Swedish e-Science Research Center.

Received 27 August 2012;
Received in revised form 25 October 2012;
Accepted 2 December 2012
Available online 7 December 2012

Keywords:
electrostatic interactions;
helical hairpin;
membrane integration;
salt bridge;
translocon

Present address: L. Martínez-Gil, Department of Microbiology, Mount Sinai School of Medicine, New York, NY, USA.

Abbreviations used:
Ct, C-terminus; ER, endoplasmic reticulum; Lep, leader peptidase; PV, poliovirus; TM, transmembrane; OST, oligosaccharyl transferase.

References

1. Wehbi, H., Gasmi-Seabrook, G., Choi, M. Y. & Deber, C. M. (2008). Positional dependence of non-native polar mutations on folding of CFTR helical hairpins. *Biochim. Biophys. Acta*, **1778**, 79–87.
2. Janovick, J. A. & Conn, P. M. (2010). Salt bridge integrates GPCR activation with protein trafficking. *Proc. Natl Acad. Sci. USA*, **107**, 4454–4458.
3. 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.
4. Gumbart, J., Chipot, C. & Schulten, K. (2011). Free energy of nascent-chain folding in the translocon. *J. Am. Chem. Soc.* **133**, 7602–7607.
5. White, S. H. & von Heijne, G. (2008). How translocons select transmembrane helices. *Annu. Rev. Biophys.* **37**, 23–42.
6. Hermansson, M., Monne, M. & von Heijne, G. (2001). Formation of helical hairpins during membrane protein integration into the endoplasmic reticulum membrane. Role of the N and C-terminal flanking regions. *J. Mol. Biol.* **313**, 1171–1179.
7. Hermansson, M. & von Heijne, G. (2003). Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane. *J. Mol. Biol.* **334**, 803–809.
8. Saaf, A., Hermansson, M. & von Heijne, G. (2000). Formation of cytoplasmic turns between two closely spaced transmembrane helices during membrane protein integration into the ER membrane. *J. Mol. Biol.* **301**, 191–197.
9. Johnson, R. M., Heslop, C. L. & Deber, C. M. (2004). Hydrophobic helical hairpins: design and packing interactions in membrane environments. *Biochemistry*, **43**, 14361–14369.
10. Baeza-Delgado, C., Marti-Renom, M. A. & Mingarro, I. (2012). Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* http://dx.doi.org/10.1007/s00249-012-0813-9.
11. Martinez-Gil, L., Bano-Polo, M., Redondo, N., Sanchez-Martinez, S., Nieva, J. L., Carrasco, L. & Mingarro, I. (2011). Membrane integration of poliovirus 2B viroporin. *J. Virol.* **85**, 11315–11324.
12. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H. *et al.* (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, **433**, 377–381.
13. Hessa, T., Meindl-Beinker, N. M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M. *et al.* (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, **450**, 1026–1030.
14. Silberstein, S. & Gilmore, R. (1996). Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **10**, 849–858.
15. Shakin-Eshleman, S. H., Spitalnik, S. L. & Kasturi, L. (1996). The amino acid at the X position of an Asn-X-ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J. Biol. Chem.* **271**, 6363–6366.
16. Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L. *et al.* (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. *J. Mol. Biol.* **387**, 1153–1164.
17. Hedin, L. E., Ojemalm, K., Bernsel, A., Hennerdal, A., Illergard, K., Enquist, K. *et al.* (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J. Mol. Biol.* **396**, 221–229.
18. Tamborero, S., Vilar, M., Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2011). Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J. Mol. Biol.* **406**, 571–582.
19. Ojemalm, K., Halling, K. K., Nilsson, I. & von Heijne, G. (2012). Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Mol. Cell*, **45**, 529–540.
20. Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. & Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. *Proc. Natl Acad. Sci. USA*, **105**, 7177–7181.
21. Nilsson, I. & von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* **268**, 5798–5801.
22. Orzaez, M., Salgado, J., Gimenez-Giner, A., Perez-Paya, E. & Mingarro, I. (2004). Influence of proline residues in transmembrane helix packing. *J. Mol. Biol.* **335**, 631–640.
23. Sauri, A., Tamborero, S., Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2009). Viral membrane protein topology is dictated by multiple determinants in its sequence. *J. Mol. Biol.* **387**, 113–128.
24. de Jong, A. S., de Mattia, F., Van Dommelen, M. M., Lanke, K., Melchers, W. J., Willems, P. H. & van Kuppeveld, F. J. (2008). Functional analysis of picornavirus 2B proteins: effects on calcium homeostasis and intracellular protein trafficking. *J. Virol.* **82**, 3782–3790.
25. Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. & Mingarro, I. (2011). N-glycosylation

efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn–Ser–Thr sequon. *Protein Sci.* **20**, 179–186.
26. Barth, P., Wallner, B. & Baker, D. (2009). Prediction of membrane protein structures with complex topologies using limited constraints. *Proc. Natl Acad. Sci. USA*, **106**, 1409–1414.
27. Monne, M., Nilsson, I., Elofsson, A. & von Heijne, G. (1999). Turns in transmembrane helices: determination of the minimal length of a "helical hairpin" and derivation of a fine-grained turn propensity scale. *J. Mol. Biol.* **293**, 807–814.
28. Hennerdal, A. & Elofsson, A. (2011). Rapid membrane protein topology prediction. *Bioinformatics*, **27**, 1322–1323.
29. Engelman, D. M. & Steitz, T. A. (1981). The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell*, **23**, 411–422.
30. Cymer, F., Veerappan, A. & Schneider, D. (2012). Transmembrane helix–helix interactions are modulated by the sequence context and by lipid bilayer properties. *Biochim. Biophys. Acta*, **1818**, 963–973.
31. Bano-Polo, M., Baeza-Delgado, C., Orzaez, M., Marti-Renom, M. A., Abad, C. & Mingarro, I. (2012). Polar/Ionizable residues in transmembrane segments: effects on helix–helix packing. *PLoS One*, **7**, e44263.
32. Buck, T. M., Wagner, J., Grund, S. & Skach, W. R. (2007). A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. *Nat. Struct. Mol. Biol.* **14**, 762–769.
33. Meindl-Beinker, N. M., Lundin, C., Nilsson, I., White, S. H. & von Heijne, G. (2006). Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep.* **7**, 1111–1116.
34. Zhang, L., Sato, Y., Hessa, T., von Heijne, G., Lee, J. K., Kodama, I. *et al.* (2007). Contribution of hydrophobic and electrostatic interactions to the membrane integration of the Shaker K+ channel voltage sensor domain. *Proc. Natl Acad. Sci. USA*, **104**, 8263–8268.
35. Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I. & von Heijne, G. (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA*, **105**, 4127–4132.
36. Monne, M., Hermansson, M. & von Heijne, G. (1999). A turn propensity scale for transmembrane helices. *J. Mol. Biol.* **288**, 141–145.
37. Nieva, J. L., Agirre, A., Nir, S. & Carrasco, L. (2003). Mechanisms of membrane permeabilization by picornavirus 2B viroporin. *FEBS Lett.* **552**, 68–73.
38. Wehbi, H., Rath, A., Glibowicka, M. & Deber, C. M. (2007). Role of the extracellular loop in the folding of a CFTR transmembrane helical hairpin. *Biochemistry*, **46**, 7099–7106.
39. Pitonzo, D., Yang, Z., Matsumura, Y., Johnson, A. E. & Skach, W. R. (2009). Sequence-specific retention and regulated integration of a nascent membrane protein by the ER Sec61 translocon. *Mol. Biol. Cell*, **20**, 685–698.
40. Sauri, A., McCormick, P. J., Johnson, A. E. & 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.
41. Sauri, A., Saksena, S., Salgado, J., Johnson, A. E. & 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.
42. Cross, B. C. & High, S. (2009). Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *J. Cell Sci.* **122**, 1768–1777.
43. Meacock, S. L., Lecomte, F. J., Crawshaw, S. G. & 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.
44. 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–L38.
45. Lundin, C., Kim, H., Nilsson, I., White, S. H. & von Heijne, G. (2008). Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for Nin/Cout and Nout/Cin transmembrane helices. *Proc. Natl Acad. Sci. USA*, **105**, 15702–15707.
46. Navarro, J. A., Genoves, A., Climent, J., Sauri, A., Martinez-Gil, L., Mingarro, I. & Pallas, V. (2006). RNA-binding properties and membrane insertion of Melon necrotic spot virus (MNSV) double gene block movement proteins. *Virology*, **356**, 57–67.
47. Martinez-Gil, L., Johnson, A. E. & Mingarro, I. (2010). Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *J. Virol.* **84**, 5520–5527.
48. Vilar, M., Sauri, A., Monne, M., Marcos, J. F., von Heijne, G., Perez-Paya, E. & Mingarro, I. (2002). Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. *J. Biol. Chem.* **277**, 23447–23452.
49. Garcia-Saez, A. J., Mingarro, I., Perez-Paya, E. & Salgado, J. (2004). Membrane-insertion fragments of Bcl-xL, Bax, and Bid. *Biochemistry*, **43**, 10930–10943.
50. Pruitt, K. D., Tatusova, T., Klimke, W. & Maglott, D. R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res.* **37**, D32–D36.

Reviewing Process

Responses to the reviewers' comments

Original comments from the reviewers are shown in **bold.**

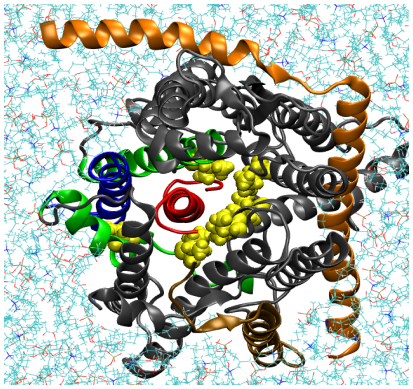
Responses to the comments are shown in plain text

Changes in the main manuscript are shown in red.

Response to the comments by the Reviewer #1.

According to the model that is based on solid experiments and presented at the end of the Discussion one has to postulate that the lysine residues in TM1 keep this domain in the translocon until TM2 has entered. Has this alreday been shown? If this is the case, the reference should be given at this position. If this has not been done, it has to be done and incorporated into the manuscript. This would strongly support the model and considerably strengthen the manuscript.

As stated by the referee, retention of TM1 in the translocon is just a putative model that might explain the mechanism underlying the establishment of electrostatic TM1-TM2 interactions prior to hairpin insertion. To our knowledge, experimental evidence demonstrating the simultaneous location of two helices inside the translocon is still missing, and to tackle such problem experimentally is beyond the scope of the present study. Nonetheless, as stated in the manuscript (page 16), unpublished molecular dynamics simulations (MDS) suggest that this possibility exists (J. Gumbart, University of Illinois at Urbana-Champaign, personal communication). In this regard, for reviewing purposes Dr. Gumbart provided us some snapshots from MDS (as the one below) that show the translocon (viewed from the cytoplasmic side, as in Fig. 6 from Gumbart & Schulten (2006) *Biophys. J.* **90**:2356-67) in a fully solvated lipid bilayer and holding two helices (blue and red in the image) within its interior:



To emphasize this fact we have included a comment in the Discussion section in line with the reviewer's suggestion:

"An interesting possibility arising from these observations is that the translocon may facilitate membrane protein integration by allowing efficient polar interactions between consecutive TM segments harbouring charged residues, and thereby stabilizing them in the nascent polypeptide. In this particular case, the lysine residues located in the first TM segment may participate in keeping this domain in the translocon until the second TM segment reaches this location. If this were the case,…"

Major points:

In Figure 1D, lane 6 the single triangle identifies the band as carrying a single N-glycosylation; this does not correlate with what is written in the text and suggested by the cartoon in 1B.

We marked this band with a single triangle because proteinase K digestion at the P1 domain renders polypeptide molecules with only one glycosylation site (G2). Nevertheless, since this can be confusing we have labeled this band with double triangle because these molecules come from the digestion of originally doubly glycosylated polypeptides, although as mentioned above, in fact this digestion product carries only one sugar moiety.

We have adapted the text (page 8) accordingly:

"…,The glycosylation site (G2) located in the beginning of the P2 domain will be modified only if this Ct domain is translocated across the membrane, while an engineered G1' site, embedded in an extended Nt sequence of 24 amino acids is always glycosylated. As previously observed, when assayed in the Lep' construct, the first hydrophobic region inserted efficiently into the membrane (up to 60% of the molecules were doubly-glycosylated, see [11] figure 4). However, synthesizing the helical hairpin in this Lep' construct in the presence of microsomal membranes the doubly-glycosylated band diminished to 34% of the molecules (Fig. 1D, lane 5), indicating a significant achievement of hairpin insertion reported as singly-glycosylated molecules (see Fig. 1C for a scheme). Furthermore, proteinase K digestions of these samples (Fig. 1D, lanes 4-6) render protected forms derived from the doubly-glycosylated molecules. Overall, …"

Most of the figures give numbers for glycosylation; the number of specif repeats should be given in the respective legend.

We have included the number of specific repeats in the respective legends (Figs. 2, 4, 5 and 6).

Figure 3 is dispensable

Although the strength of this study lies in the experimental approach, computational modeling displayed in Figure 3 has partially guided our experimental designs. In order to reinforce this fact, we have generated new models using a more restrictive cutoff for the interaction between the polar groups and updated the figure and the text accordingly. Of course in these new conditions the fractions are lower but the preference for Asp74 to interact with Lys42 and Lys46 over Lys 39 is even more

pronounced. In addition, to make the figure more informative we have now updated both the caption and legend (page 21). Changes in the body text (page 10) and figure legend are as follow:

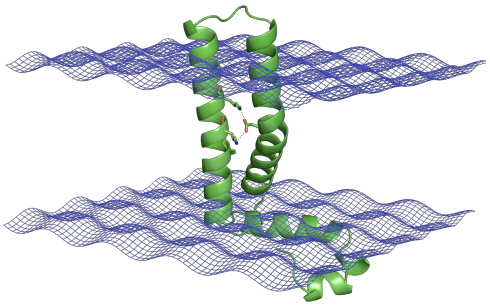
"These models clearly predicted a preference for Lys46 and Lys42 to interact with Asp74. Among the five largest structural clusters obtained, as compared to Lys39, interactions of Lys42 and Lys46 with Asp74 are 53 and 75 times more frequent respectively (Figure 3)."

Figure 3. Observed interactions of Asp74 among low energy models generated by the Rosetta-Membrane protocol (see Methods). Clearly, it is energetically favorable for the hairpin to hide the polar groups of the Asp and Lys residues. In particular the polar group of Asp74 can be shielded from the hydrophobic lipid environment by interactions with one of the three Lys in the first helix. The data analysis shows that interactions between Asp74 and either Lys42 or Lys46 are found much more frequently than interactions between Asp74 and Lys39, which are almost never observed. The inset shows one of the models where interactions are seen between Asp74 and both Lys42 and Lys46.

Minor points:

The graphical abstract is not very informative

We include a new graphical abstract, which we believe that is more informative than our previous one. In this new image (below) we show the model of full-length PV 2B in the context of a membrane and highlight the interactions between the charged residues found in our work.



On page 5, third paragraph the authors should refer to plasma membrane proteins (their statement is not true for mitochondrail membrane proteins)

We have referred for plasma membrane proteins in the current version (page 4).

There is something missing from Figure 3

We have improved Figure 3 caption, see above.

Semantics:

Difficult is an adjective and not a verb, replace by make difficult or complicate

We have replaced difficult by complicate in the Abstract.

Cytoplasm includes all organelles except for the nucleus; in the text and the figures cytosplasm has to be replaced by cytosol

We have replaced cytoplasm by cytosol both in the body text and figures.

mRNA is translated, proteins are not translated, they are synthesized; please ammend manuscript accordingly

We have amended this feature throughout the manuscript.

Response to the comments by the Reviewer #2.

1. Pages 8-11 and Fig.1. The data give rise to ambiguous interpretation. Fig 1B and 1C can be interpreted as no helix hairpin insertion at all either in the inner and outward-facing orientation. A similar issue could be true of the data in 1C and 1D.

We agree with the reviewer that the results in Fig. 1 as displayed in the previous version of the m/s could give rise to ambiguous interpretations. Therefore, to eliminate ambiguity, we have added supplementary information to the new version of this figure. Using the Lep system, isolated TM segments are translocated across the microsomal membranes (double glycosylation amounts above 95%). When the system is challenged with the hairpin sequence (including both TM segments) a significant level of insertion is observed (single glycosylation, 31%). In this Lep set-up the presence of singly-glycosylated forms can only be explained by hairpin insertion. Since the topology of the hairpin in this system is opposed to the one suggested for the 2B protein, we sought to use the Lep-derived (Lep') system (also described in reference 11, fig 4). In this system, 2B-HR1 is inserted with the proper topology, which is reflected by the prevalence of doubly-glycosylated forms (60 %). Addition of the HR2 helix to test the whole hairpin results in the increase of singly-glycosylated forms (from <40% to 66%). The latter observation is consistent with insertion of the HR2, i.e., with insertion of the full hairpin. Although, HR2-induced blocking of HR1 insertion (i.e., precluding hairpin insertion) could be invoked to explain these results, we do not support that interpretation because our data in the earlier and the present work strongly demonstrate that 2B is an integral membrane protein (see Figs. 5-7 from reference 11, and Figs. 2-6 in the present manuscript). The text has been adapted (page 8) to include these comments, see above our response to the first Major point from **reviewer #1**.

A glycosylation site may need to be inserted into the linker region between the two putative helices. to define insertion more rigorously.

Although an interesting approach, we discarded the possibility of inserting a glycosylation site into the linker region because, as demonstrated previously and

mentioned in the text (page 7, references 20 and 21), efficient glycosylation occurs when the acceptor Asn is ≈12-14 residues away from the membrane. This means that in the present case a foreign sequence of above 20-25 residues (the native linker contain only 5 residues) should be need it to introduce a new glycosylation site, which will probably strongly influence the topology of the synthesized molecule. This is, in our opinion, very likely in the current case because as demonstrated later on in the manuscript the residues included in the linker region play an important role in hairpin insertion (see Fig. 6).

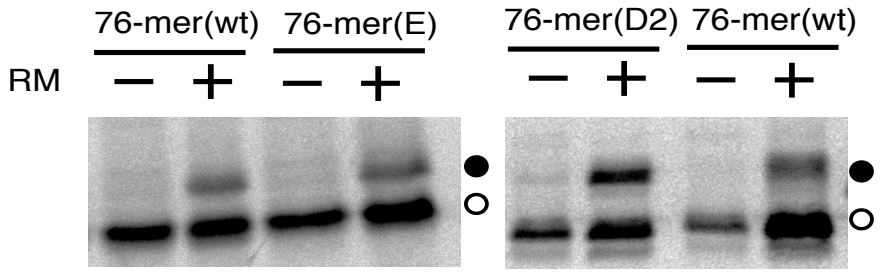
All in all, we believe that Fig. 1 represents an interesting starting point for the current work.

2. The data in Fig 2. are not convincing. There is a significant degree of glycosylation in the wild type, indicative of a failure to form the helical hairpin in the membrane. Again there is considerable non-glycosylated material in the K mutants.

In the transcription/translation system in the presence of microsomal membranes the glycosylation efficiency is commonly the result of an equilibrium between molecules with different membrane disposition. The data included in Fig. 2 demonstrate that for the wild-type sequence the above mentioned equilibrium is balanced towards the inserted hairpin, whilst in the case of the constructs that include lysine residues replacing the native Asp74 the more predominant form is the one with a translocated C-terminus. To clarify this point, we have included schemes of these equilibria as part of panel B in the revised version and adapted the figure legend.

No gel data are provided for the D mutants and the RM control (1st lane) is unexplained-what is the strong band in the glycosylated position and why not in other control lanes? As with 1 above the system as designed is suggestive rather than definitive.

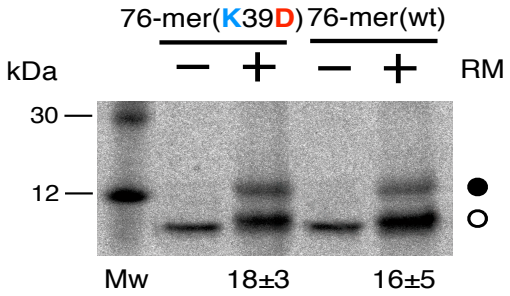
We did not include the results obtained for the acidic mutants as a gel panel to avoid overloaded images. We think that the plot representation in panel C gives a more accurate comparison for those samples with closer glycosylation efficiencies like the ones observed for the acidic mutants. Just as an example, we include for reviewing purposes two representative SDS-PAGE autoradiography images for the mutants mentioned. The glycosylation levels observed in these captions was used to calculate the averaged value together with the data obtained from two other independent experiments:



Concerning the 1st lane in panel B, this is not a RM control. This lane contains the radioactive molecular weight (Mw) markers used as a reference for the size of the truncate molecules. Since our previous version resulted somehow confusing we have introduced the appropriate modifications in Fig. 2 to clarify this item (see new Fig. 2) and adapted the figure legend.

3. Fig4. This illustrates a possible long range/dynamic interaction between position 74 and positions 42 and/or 46. The suggestion would have been more convincing if similar experiments with position 38 and 39 were shown to support specificity in the observations.

The design of our mutations was guided by computational modeling, which pointed out that Lys39 only interacts with Asp74 in less than 1% of the structural models generated using the Rosetta-Membrane protocol (see Fig. 3). According to these predictions it is not very likely that Lys39 contributes significantly to stabilize hairpin insertion. Nevertheless, we have produced the K39D mutant and performed the insertion/glycosylation experiments. We have found glycosylation levels for this mutant that were indistinguishable to the ones observed for the wild-type sequence (18±3 *vs* 16±5 glycosylation percentage respectively, see figure below), suggesting that Lys39 does not participate directly in the hairpin insertion process:



To include these results, we have adapted the text in the current version (page 10) as follows:

"In a series of experiments performed in hairpin truncations (76 residues long polypeptides, 76-mers) we separately replaced each lysine residues by an aspartic acid residue. As expected, K39D mutation exhibited a glycosylation level indistinguishable from the wild-type sequence (data not shown), indicating that it does not perturb hairpin insertion and confirming our computational predictions. On the contrary, as shown in Fig. 4A,…"

Moreover, the results do not appear to be consistent in that the D46K74 mutant seems to have increased rather than reduced glycosylation.

For the wild-type protein, Asp74 can both interact with Lys46 and Lys42, whilst for the mutant mentioned by the referee (Asp46, Lys74) we could partially restore the pair charge interaction. Nonetheless, we note that at position 42 there is a Lys residue in this construct, which could be not so well suited for interacting with Lys74. Although this is just a speculation it is consistent with the slightly higher level of glycosylation for this mutant (19±6) compared to the wild-type sequence (16±5).

4. The data in Fig. 5 are confusing. The FL construct contains positively charged residues at the C-terminus and on the basis of previous data/arguments (Fig.4), it should have been significantly glycosylated since there were no D residues in the peptide. It was not-perhaps because of the +ve inside rule. Addition of asp residues to the first helix increased glycosylation when it would have been expected to reduce it?

The positively charged residues (82-84) are not included in the TM region, they are actually located in the flanking region. As proposed by the positive inside rule those residues could interact with the head groups of the phospholipids stabilizing protein topology. In fact, addition of Asp residues preceding the first TM region would perturb (according to the positive inside rule) the membrane disposition of the full-length constructs.

Nevertheless, in order to explicit the flanking (extra-membrane) location of residues 82-84, we have modified the main text (page 14) and the legend for this figure as follows:

"… . Interestingly, it has been previously demonstrated that extramembranous positively charged residues placed near the cytosolic end of TM segments in membrane proteins promote membrane insertion of precedent hydrophobic helices [17; 31]. Inspection of PV 2B sequence reveals the presence of three positively charged flanking residues adjacent to the helical hairpin C-terminal end (see Fig. 5, top), which could then contribute to the higher integration efficiency for the full-length protein compared to truncated molecules where these flanking positively charged residues are absent. …"

"… . Structural organizations of full-length PV 2B constructs are shown on top, with the flanking basic residues Arg82, Lys83 and Lys84 highlighted in blue. Bottom:…"

5. The results with the linker region look more convincing and are consistent with observations made elsewhere

We agree with the reviewer, the results observed for the turn region mutants cement the observations made in the rest of the manuscript, emphasizing the overall message of the present work. We have stressed this message in the Discussion section as follows:

"We also find that the contribution of turn residues can be combined with the interaction between charged residues appropriately located to stabilize the helical hairpin conformation by inter-helical hydrogen bonding, since the higher levels of glycosylation have been obtained when perturbed both turn and charged residues at the same time (Fig. 6B). Similarly, structural perturbations introduced by extracellular loop mutations in a cystic fibrosis TM conductance regulator (CFTR) appeared to affect hairpin conformation to a lesser extent in the context of more stable TM helix-helix interactions [36], highlighting the interplay between helix-helix interactions and turn contribution in modulating hairpin formation."

Chapter 4:

Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing

OPEN ACCESS Freely available online

Article 4

PLOS ONE

Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing

Manuel Bañó-Polo[1], Carlos Baeza-Delgado[1], Mar Orzáez[2], Marc A. Marti-Renom[3,4], Concepción Abad[1], Ismael Mingarro[1]*

1 Departament de Bioquímica i Biologia Molecular, Universitat de València, Burjassot, Spain, **2** Centro de Investigación Príncipe Felipe, Valencia, Spain, **3** Genome Biology Group, Structural Genomics Team, Centre Nacional d'Anàlisi Genòmic, Barcelona, Spain, **4** Structural Genomics Group, Center for Genomic Regulation, Barcelona, Spain

Abstract

The vast majority of membrane proteins are anchored to biological membranes through hydrophobic α-helices. Sequence analysis of high-resolution membrane protein structures show that ionizable amino acid residues are present in transmembrane (TM) helices, often with a functional and/or structural role. Here, using as scaffold the hydrophobic TM domain of the model membrane protein glycophorin A (GpA), we address the consequences of replacing specific residues by ionizable amino acids on TM helix insertion and packing, both in detergent micelles and in biological membranes. Our findings demonstrate that ionizable residues are stably inserted in hydrophobic environments, and tolerated in the dimerization process when oriented toward the lipid face, emphasizing the complexity of protein-lipid interactions in biological membranes.

Citation: Bañó-Polo M, Baeza-Delgado C, Orzáez M, Marti-Renom MA, Abad C, et al. (2012) Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing. PLoS ONE 7(9): e44263. doi:10.1371/journal.pone.0044263

Editor: Peter Butko, Nagoya University, Japan

Received April 25, 2012; **Accepted** July 31, 2012; **Published** September 12, 2012

Funding: This work was supported by grants BFU2009-08401 (to IM) and BFU2010-19310 (to MAM-R) from the Spanish Ministry of Science and Innovation (MICINN, ERDF supported by the European Union), as well as by PROMETEO/2010/005 (to IM) from the Generalitat Valenciana. MB-P and CB-D were recipients of FPU and FPI predoctoral fellowships from the MICINN, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

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

Introduction

The vast majority of membrane proteins are anchored to biological membranes through hydrophobic α-helices. These transmembrane (TM) α-helices, rather than serving solely as featureless hydrophobic stretches required for anchorage of proteins in membranes, have structural and/or functional roles well beyond this canonical capacity. In fact, the folding and assembly of membrane proteins rely in part on interacting TM helices, which was conceptualized as a two-stage process [1]. In the first stage, TM helices are inserted into the membrane by the translocon. The driving force for this process derives primarily from the transfer of hydrophobic side chains from the aqueous channel of the translocon to the apolar region of the bilayer [2]. In the second stage, the protein attains its native tertiary structure through the packing of its TM helices. In the apolar environment of the membrane core, van der Waals packing, hydrogen bonding and ionic interactions are the dominant contributors to TM helix packing.

Sequence analysis of high-resolution membrane protein structures show that ionizable amino acid residues are present in TM helices, although at a low frequency level [3]. Insertion of these residues through the translocon has been proved to be feasible thanks to the overall hydrophobicity of the TM segment [4] and depending on their position along the hydrophobic region [5]. In many cases, ionizable residues are involved in TM helix packing [6,7,8]. Likely, hydrogen bonding [6,7] or salt-bridge [9] formation with other membrane-spanning hydrophilic residues drives these interactions, while at the same time, reduces the unfavorable energetics of inserting polar or ionizable residues into the hydrophobic membrane core.

Homo-oligomeric membrane proteins provide attractive systems for the study of TM helix packing because of their symmetry and relative simplicity. These model systems can serve as an excellent starting point to understand the structural dynamics and folding pathways of larger membrane proteins. One of the best-suited models of membrane protein that oligomerizes (more specifically, dimerizes) through non-covalent interactions of its TM α-helix is undoubtedly glycophorin A (GpA) [10,11]. The wide use of this protein as a model membrane protein is partially based on its intrinsic simplicity, since the free energy decrease associated with TM helix-helix interactions is enough to confer detergent resistant dimerization to the protein. Thus, those factors that could affect or modify the dimerization process can be analyzed using sodium dodecyl sulfate (SDS)-PAGE. The GpA homodimer, defines a dimerization interface that has been extensively studied by diverse techniques such as saturation mutagenesis [12] and alanine-insertion scanning [13] in SDS micelles, solution NMR in dodecyl phosphocholine micelles [14] and solid-state NMR in lipid membranes [15]. The output of these studies describes a dimerization motif in the TM segment composed of seven residues, L^{75}IxxGVxxGVxxT^{87}, which is responsible for the dimerization process. More recently, using proline-scanning mutagenesis it was demonstrated that Leu75 is not so cleanly involved in the packing process [16], focusing the interaction on the central G^{79}VxxGVxxT^{87} motif, which includes the widely proved framework for TM helix association, GxxxG [17,18].

Nevertheless, the sequence context highly determines the thermodynamic stability of GxxxG-mediated TM helix-helix interactions (recently reviewed [19]).

In the present study, we have analyzed the distribution of ionizable (Asp, Glu, Lys and Arg) amino acid residues in TM segments from high-resolution membrane protein structures, which have to energetically accommodate into the highly hydrophobic core of biological membranes by interacting favorably with its local environment. Then, we address the consequences of replacing specific residues by ionizable amino acids along the hydrophobic region of the GpA TM domain on the dimerization of this model membrane protein, both in detergent micelles and in biological membranes. Our findings demonstrate that ionizable residues are stably inserted in hydrophobic environments, and tolerated in the dimerization process when oriented toward the lipid face, emphasizing the complexity of protein-lipids interactions in biological membranes.

Results and Discussion

Ionizable amino acid residues in TM α-helices

TM helices of lengths between 17 and 38 residues were selected from the MPTOPO database [20], which included helical segments that do completely span the hydrophobic core of the membrane. TM helices shorter than 17 residues as well as larger than 38 residues were excluded since they may not cross entirely the membrane or may contain segments parallel to the membrane [3], respectively.

As expected, ionizable residues (Asp, Glu, Lys, and Arg) are present at a low frequency level. All together, these residues constitute only 6.6% of the residues within TM helices. Despite their lower presence, strongly polar residues are evolutionary conserved in TM proteins, which can be partially explained by their tendency to be buried in the protein interior and also in many cases due to their direct involvement in the function of the protein [21,22]. Among the 792 TM helices included in our database, 366 helices (46.2%) contained at least one ionizable residue within the hydrophobic region (that is, the central 19 amino acid residues). A summary of the statistics is presented in Figure 1. Furthermore, 96 TM helices contained at least one acidic plus one basic residue in their sequence, and 20 of these helices present oppositely charged residues with the appropriate periodicity (*i*, *i*+4) to form intrahelical charge pairs. To gain more detailed insight into the structural role of these ionizable residues within the membrane core, we analyzed the environment of all these 20 helices. Approximately half of the ionizable residues (51%) found in these helices are buried in the protein interior, but the rest are partly exposed to the lipid face. Some of these lipid facing ionizable residues are located in pairs at the appropriate distance to form a salt-bridge, as in the sarcoplasmic/endoplasmic reticulum calcium ATPase 1 protein (Fig. 2).

Effects on SDS-resistant TM helix packing

Next, we investigated the effect of strongly polar residues in TM helix packing using the GpA TM segment as a model (scaffold) segment. Initial polar mutations (T87D, T87K, I91D, and I91K) made on residues located at the helix-helix interface (Fig. 3A) abolished dimerization (Fig. 3B). Furthermore, it has been reported that T87S (which retains the side chain γ oxygen) permits dimer formation both in SDS micelles [23] and in *E. coli* membranes [24], whereas a bulkier hydroxylated side chain (T87Y) is strongly disruptive (Fig. 3B). However, point mutations corresponding to replacements of nonpolar residues located at the lipid-facing interface (Fig. 3A) by ionizable residues gave rise to a more tolerated response (Fig. 3B). When Ile85 was substituted by ionizable side-chain residues, either negatively charged (I85D) or positively charged (I85K and I85R), the dimerization level was similar to native GpA TM sequence as shown under SDS-PAGE analysis (Fig. 3C, compare lanes 2, 3 and 4 to lane 1). It is commonly assumed that single ionizable residues should exist in their uncharged form within membrane-spanning helices [25]. In fact, the pK_a values observed for Asp residues in hydrophobic helices were somewhat elevated (5–8.5) relative to those for Asp residues in solution [26]. Furthermore, the replacement of Leu89 by basic residues (L89K and L89R) had almost no effect, while its substitution by an acidic residue (L89D) abolished dimerization (Fig. 3B and 3C). The opposing consequences observed for Leu89 mutants can be explained taking into account the nature of the SDS-micelles used in these experimental conditions. These results suggest that L89D mutation alters the interaction of the protein with the negatively charged detergent micelle, possibly resulting in a structure that differs from a 'transmicellar' α-helix due to helix distortions and interaction with the polar micelle surface. This effect was not observed when the Asp residue was located in a more central position (I85D), where its carboxylate should be located away from the negatively charged sulfate groups of the SDS molecules. In this regard, the capacity of SDS to respond to such nuance of sequence in terms of SDS solvation of TM segments within protein-SDS detergent complexes has been proved to be highly sequence (position) dependent [27]. Nevertheless, the comparable electrophoretic migration observed for I85D and L89D (Fig. 3C) suggests that the monomers associate with SDS quite similarly. To identify the helix interface responsible of dimer formation in the Leu89 mutants, we designed double mutants that contained a non-polar highly disruptive mutation (G83L). Gly83 has been proved to be extremely sensitive, since all mutations tested disrupted the dimer completely [12]. As

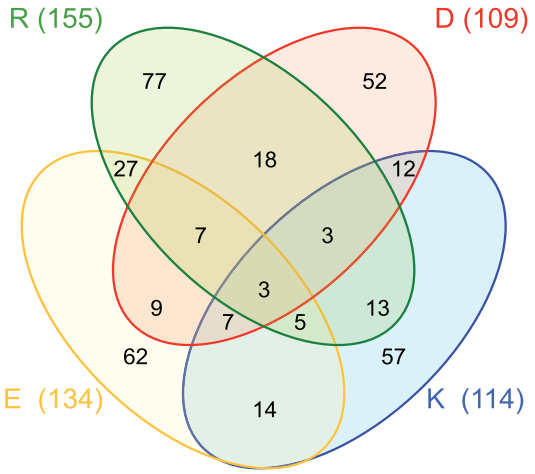


Figure 1. Venn diagram of TM segments (the central 19 residues) containing charged residues: Asp (D, red), Lys (K, blue), Glu (E, yellow) and Arg (R, green). The value in parenthesis is the total TM helices that contain at least one of such residues. The values inside the ellipses indicate the number of TM helices in each combination of these four amino acids. For example, there are 57 TM helices with only Lys as a charged residue, 12 helices with only Lys and Asp, 7 helices with Lys, Asp and Glu, and 3 helices with all four ionizable residues.
doi:10.1371/journal.pone.0044263.g001

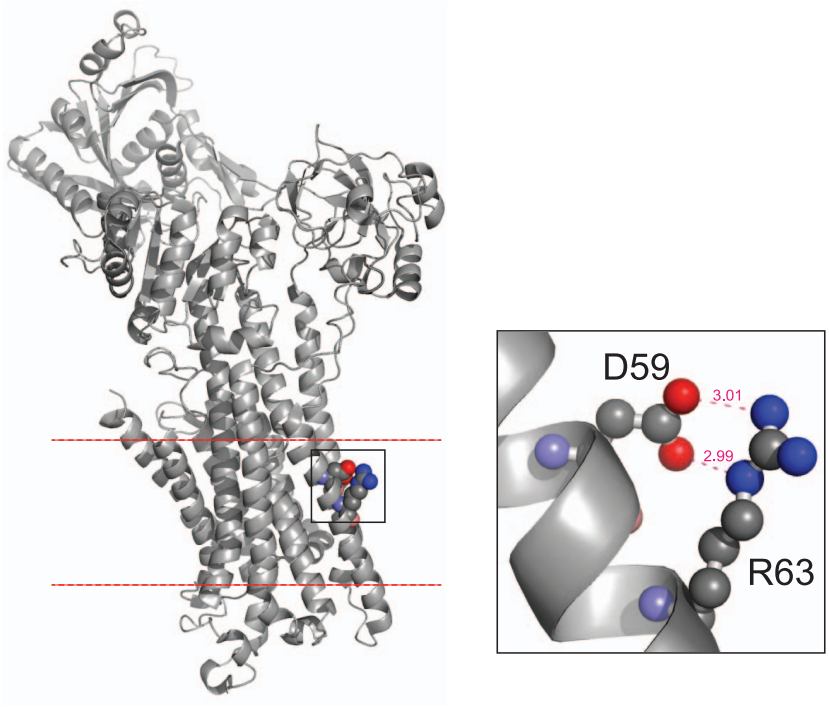


Figure 2. Structure of the calcium ATPase 1. Left panel: cartoon representation of sarcoplasmic/endoplasmic reticulum calcium ATPase 1 (PDB ID: 1SU4) with cytosolic domain in the up side and transmembrane aspartic 59 and arginine 63 residues in spheres representation (C atom gray, O atom red and N atom blue). Membrane boundaries (dashed red lines) were obtained from the *PPM Server* [44]. **Right panel:** zoom view centered on the salt bridge between Asp59 and Arg63, dashed pink lines indicate O to N atom distances.
doi:10.1371/journal.pone.0044263.g002

shown in Fig. 3B, G83L mutant did not form any detectable dimer, and both double mutant proteins (G83L/L89K and G83L/L89D) containing this mutation did not dimerized, suggesting that the lysine residue introduced was not participating in the dimerization process, instead, the native dimerization motif is responsible of helix-helix interaction.

Given the 3.6-residue periodicity of an ideal α-helix, intrahelical charge pairs would be expected for (*i*, *i*+4) Lys-Asp pairs. To further assess if intrahelical charge pair formation can be tolerated in dimerizing TM sequences, we performed a double mutation combining two strongly dimerizing sequences (I85D/L89K), which only reduced dimerization by about 50% compared to the wild-type sequence (Fig. 3B). Similarly, I85K/L89K mutant retained the same level of dimerization, likely favored by a beneficial SDS solvation effect on the lysine residues. On the contrary, when oppositely charged residues were located at the TM-interacting interface (T87D/I91K) dimerization was abrogated (Fig. 3B). Furthermore, when charge pairs include L89D mutation although facing the lipids, as for I85K/L89D, we found no evidence for dimer formation (Fig. 3B). These results suggest that charge pairs are tolerated only when located at the non-interacting interface, but solely at specific positions.

Recent mutational analysis of strongly self-interacting TM segments demonstrated that basic and acidic residues located at the helix-interacting interface participate in homotypic interactions [25]. In this case, basic and acidic residues spaced (*i*, *i*+1) and (*i*, *i*+2) contribute to the interaction of model TM segments. To test this idea in the GpA sequence, we designed two mutants with appropriately spaced basic and acidic residues (L89D/I91K and L90D/I91K), and no dimeric forms were observed in any of these proteins.

In light of our experiments in SDS micelles, it can be concluded that nonpolar to ionizable substitutions away from the dimer interface (lipid facing) in combination with N-terminal native GpA dimerization motif (including GxxxG sequence) does not perturb the dimerization process, while similar mutations positioned at the helix-interacting interface strongly compromise dimer formation.

Effects on insertion and packing into biological membranes

To test the molecular effect of the ionizable residues in biological membranes we used a glycosylation mapping technique to measure changes in the insertion capacity of the GpA TM domain after introduction of ionizable residues at the more tolerant positions in terms of TM packing. The glycosylation mapping technique has been used previously to investigate the membrane insertion level of hydrophobic regions and to systematically examine the effects of individual residues on their position in the membrane [16,28,29]. The method is based on the observation that the endoplasmic reticulum (ER) enzyme oligosaccharide transferase (OST) can only transfer a sugar moiety to Asn-X-Thr/Ser acceptor sites when they are oriented toward the lumen of the ER membrane. To assess the effect of the presence of ionizable residues on the GpA TM segment insertion into

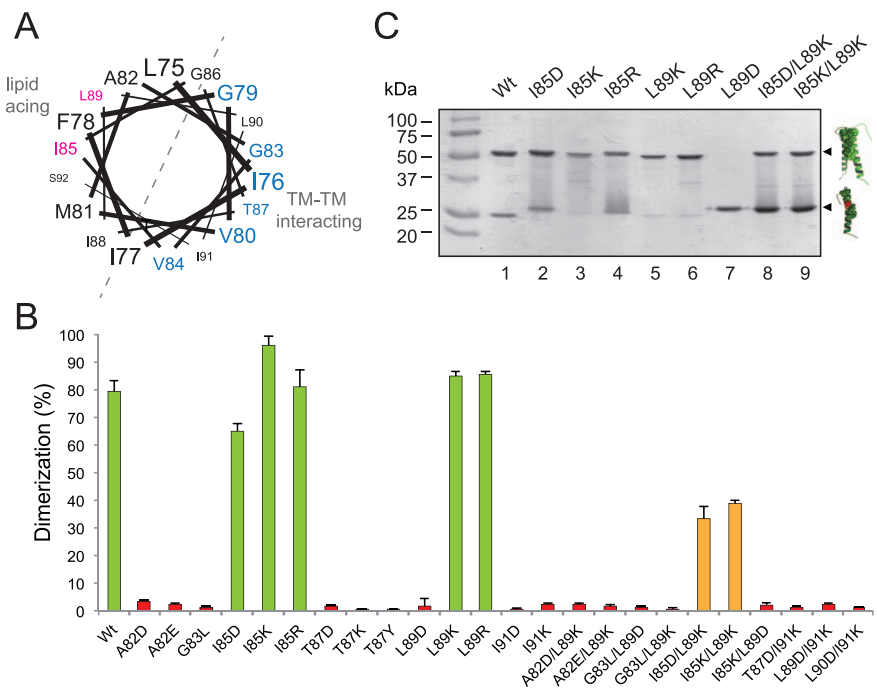


Figure 3. Dimerization in SDS micelles. (A) Helical wheel projection of GpA TM sequence. The residues associated with dimer formation as defined by Orzaez et al. [16] are shown in blue. Non-interacting residues susceptible of ionizable residue substitution are shown in magenta. (B) Green colored bars denoted dimerization levels similar to wild-type sequence. Bars for intermediate dimerization levels (≈40%) are colored orange. Red colored bars denote non-dimerizing sequences (dimerization<3%). (C) SDS-PAGE analysis of GpA mutants. Chimeric proteins were purified in the presence of SDS and analyzed by PAGE. Positions of the monomer and dimer of the chimeras are marked on the right as single and double helices, respectively.
doi:10.1371/journal.pone.0044263.g003

biological membranes, we located this hydrophobic sequence (Fig. 4A) in place of the second TM fragment of the well-characterized *Escherichia coli* inner membrane protein leader peptidase (Lep). Although of bacterial origin, Lep integrates efficiently into dog pancreas microsomes with the same topology as in *E. coli* [30] (*i.e.*, with both the N- and C-termini exposed to the luminal side of the ER membrane) and the presence of its first TM segment together with the cytoplasmic P1 domain (Figure 4B) is sufficient for proper targeting of chimeric proteins to the eukaryotic membrane [30,31]. An engineered glycosylation site placed at the C-terminal P2 domain is glycosylated efficiently upon correct insertion into the microsomal membrane (Fig. 4B), serving as a reporter to distinguish between a lumenal (glycosylated) and a cytoplasmic (unglycosylated) location. Glycosylation of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep expressed in the absence of microsomes. The efficiency of glycosylation of Lep under standard conditions is 80–90% [31,32]. The strength of the Lep system is that it provides a comparative scale for the energetic cost of inserting a broad range of model and actual TM sequences into biological membranes, closely mimicking the *in vivo* situation.

The wild-type sequence of GpA TM segment efficiently inserts into the ER-derived microsomal membranes, while I85D mutation severely diminished membrane insertion capacity (Fig. 4C). On the contrary, L89K mutation allowed efficient insertion (Fig. 4C, lane 6). The different effect observed for these two mutants can be explained by differences in amino acid side chain size and the position of the residue in relation to the midpoint of the TM sequence (Fig. 4A). Hence, in the case of L89K, the longer side chain of this cationic amino acid and its proximity to the membrane interface compared to I85D may allow the hydrophilic moiety of the lysine residue to snorkel, that is, to approach its ε-amino group toward the interfacial and aqueous region, close to the negatively charged phospholipid head groups. Next, a construct with an Asp-Lys pair at the same positions (double mutant I85D/L89K) was glycosylated somewhat more efficiently than the I85D construct (Fig. 4C, lanes 4 and 8), supporting the idea that an intrahelical salt-bridge or hydrogen bond interactions between Lys and Asp side chains located on the same face of a TM helix can facilitate its insertion into biological membranes by reducing the free energy of membrane partitioning, as previously suggested in a similar system [9]. Furthermore, the predicted insertion frequencies from the biological hydrophobicity scale [2,5] for these mutants using the ΔG Prediction Server v1.0 (http://dgpred.cbr.su.se/) are shown in Table 1. In this algorithm, the predicted insertion frequency comes from the apparent free-energy difference (ΔG_{app}) from insertion into ER membranes. Since very low and very high insertion efficiencies cannot be accurately measured, ΔG_{app} values outside the interval ±1.5 kcal/mol are only qualitative. The positive value of ΔG_{app} predicted

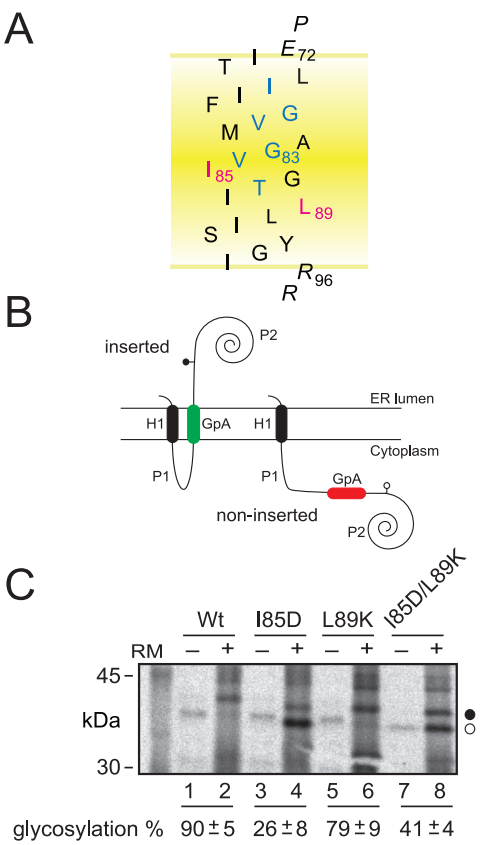


Figure 4. Insertion of GpA-derived segments into microsomal membranes. (A) Model of the GpA TM wild-type sequence. GpA residues involved in dimer formation are blue, the hydrophobic residues replaced to ionizable residues are magenta, and flanking residues are shown in italic. (B) Membrane topology of Lep chimeras. The second TM segment of Lep was replaced by the GpA TM amino acid sequence (gray). The glycosylation acceptor site located in the beginning of the P2 domain will be modified only if GpA-derived TM sequence inserts into the membrane. (C) *In vitro* translation in the presence of rough microsomal membranes (RM). Proper insertion of the GpA-derived TM sequences results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of the proteins expressed in the absence of microsomes. Bands of nonglycosylated and glycosylated proteins are indicated by white and black dots, respectively. Average ± s.d. of glycosylation results from four independent experiments are shown at the bottom.
doi:10.1371/journal.pone.0044263.g004

that the tested-sequence is not TM. The high negative value for the GpA wild-type sequence agrees with our experimentally measured glycosylation data showing the highest insertion efficiency. A closer analysis of the output data highlighted I85D mutation as precluding TM disposition. Hence, replacing Ile85 with aspartic acid reduced ΔG_{app} by almost 2 kcal/mol (Table 1), which correlates with our lowest glycosylation efficiency. However, replacing Leu89 with lysine has a lower energy cost (ΔG_{app} close to 0) that is reflected by a higher insertion level (Fig. 4C). Finally, the double mutant I85D/L89K results in the highest predicted penalty for TM disposition, whereas experimentally we find no evidence that GpA TM segment is significantly compromised by the presence of two poplar/ionizable residues. Such phenomena points towards an intra-helical interaction between the ionizable residues and should be taken into account to improve TM prediction algorithms.

Finally, the effect of ionizable residues in TM packing in bacterial cytoplasmic membranes was assessed using the ToxCAT assay [33]. This assay uses a chimeric construct composed of the ToxR N-terminal transcriptional activation domain [34] fused to the GpA TM segment and a C-terminal maltose binding protein (MBP) domain (Fig. 5A). TM-mediated dimerization of the chimera in the *E. coli* inner membrane results in transcriptional activation of a reporter gene encoding chloramphenicol acetyltransferase (CAT), with the level of CAT protein expression indicating the strength/intensity of TM helix-helix interactions. After transformation of these ToxCAT constructs into *E. coli* NT326 cells, we tested the ability of the wild-type and mutant fusion proteins carrying ionizable residues to complement the *malE* phenotype of the NT326 strain by growing each construct on plates containing maltose as the sole carbon source. Cells containing a construct that lack a TM segment do not grow (pccKAN), but the wild-type and all point mutants support growth on maltose (Fig. 5B), indicating that the MBP domains of these chimeric proteins are properly targeted to the periplasm of the NT326 cells. Consequently, the expected topology (Fig. 5A) is being achieved by these proteins, in agreement with GpA wild-type and point mutants in ToxR [35] and ToxCAT [33] assays. Dimerization of wild-type and mutant sequences carrying ionizable residues was assessed along with a GpA point mutant (G83I) that disrupts homodimerization as negative control. The I85D mutant was found to dimerize in this system to about 35% of the level shown by wild-type GpA (Fig. 5C). Interestingly, L89D mutant, which precludes dimer formation in the presence of SDS micelles (Fig. 3C), appears to retain some dimerization capacity (21±4%, normalized dimerization), which highlights the influence of the specific lipid environment during the assembly of TM segments [36]. Nevertheless, differences in TM segment length and flanking residues sequences (see Fig. S1) may alter the dimerization process in the two systems, which are difficult to rationalize. Mutation of Leu89 to lysine (L89K) had a smaller effect on TM dimerization, and double mutant I85D/L89K still retained some dimerization capacity (Fig. 5C). In agreement with these data, recent molecular dynamics simulations suggested that a lysine residue outside the contact interface could exert a significant influence on TM helix association affinity of the bacteriophage M13 major coat protein because the extent of their burial in the membrane could be different in monomers and dimers [37]. Together, our data indicate that the presence of ionizable residues does not preclude membrane insertion and allows dimer formation in bacterial cells.

Conclusions

Ionizable amino acid residues are functionally and/or structurally important residues in membrane proteins. Therefore, although the insertion of such residues into the membrane hydrophobic core may be energetically unfavourable, there is often a functional and/or structural necessity to accommodate them. In the light of our experiments it can be concluded that nonpolar to ionizable point substitutions at specific positions away from the dimer interface ('lipid facing') in combination with a N-terminal GxxxG motif does not preclude neither the dimerization process nor TM helix insertion, while point mutations of nonpolar (or polar nonionizable) to ionizable residues in the 'helix facing',

Table 1. Thermodynamic cost of GpA-derived TM segments integration.

GpA-derived region	ΔG_{pred}	Glycosylation % (measured)	Sequence
Wt	−1.646	90±5	ITLIIFGVMAGVIGTILLISYGI
I85D	+ 0.413	26±8	ITLIIFGVMAGV**D**GTILLISYGI
L89K	+ 0.112	79±9	ITLIIFGVMAGVIGTI**K**LISYGI
I85D/L89K	+ 2.561	41±4	ITLIIFGVMAGV**D**GTI**K**LISYGI

The predicted (ΔG_{pred}) energetic cost in kcal/mol of inserting versions of the GpA TM spanning region estimated using the biological hydrophobicity scale [2,5] are provided solely for the basis of comparison. Negative ΔG_{pred} values are indicative of TM disposition, while positive values indicate non-TM disposition. Mutated residues at positions 85 and 89 are shown in bold.
doi:10.1371/journal.pone.0044263.t001

e.g. I91D/E, I91K/R, or T87D strongly compromise dimer formation. These notions need to be considered if we are to develop a predictive understanding of TM helix interactions in membrane proteins.

Materials and Methods

Helix data set

All α-helical membrane proteins deposited in the MPTOPO database (last updated on January 19th, 2010) [20], and thus with known membrane insertion topology, were selected. The initial set was further filtered by: (i) removing any entry of unknown structure as based on the MPTOPO entry classification (*i.e.*, keeping only entries described as "3D_helix" and "1D_helix"); and (ii) removing redundant pairs at 80% sequence identity by applying the *cd-hit* program [38]. The final data set of TM helices contained 170 non-redundant structures, 837 TM helices, and 20,079 amino acids. Furthermore, to properly analyze the amino acid propensities in single membrane spanning TM helices, we discarded any helix shorter than 17 amino acids or larger than 38 amino acids. The resulting TM data subset contained 792 TM helices, and 19,356 amino acids.

Plasmid constructs

Construction of the plasmids encoding the His-tagged chimeric proteins (SN/GpA) have been described [13,39]. Mutations at the TM fragment of GpA were obtained by site-directed mutagenesis using the QuikChange site directed mutagenesis kit (Stratagene, La Jolla, California). Introduction of the TM segment from GpA into the Lep sequence was described elsewhere [16]. The ToxCAT vector pccKAN, and the derivatives carrying the TM domain of GpA (pccGpA) and a disruptive GpA mutant (pccGpA-G83I) fused to the ToxR transcription activator and to maltose-binding protein (MBP) were described previously [33]. All mutants were confirmed by DNA sequencing.

Protein expression and purification

Overexpression and purification of His-tagged SN/GpA constructs from transformed *Escherichia coli* BL21 (DE3) cells was performed as described [40]. *In vitro* transcription/translation of Lep-derived constructs was done in the presence of reticulyte lysate and [^{35}S]-labeled amino acids as described [16].

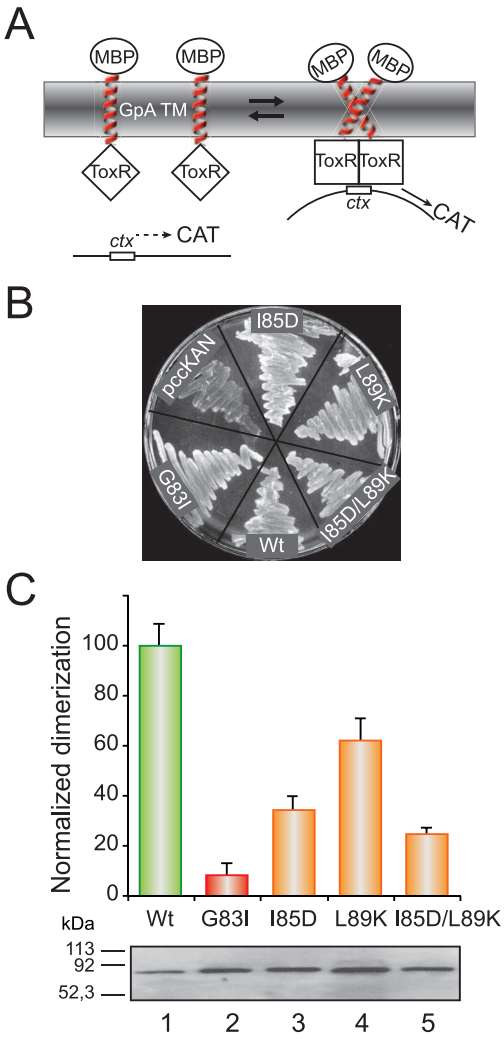


Figure 5. Dimerization in *E. coli* membranes. (A) Schematic representation of the ToxCAT assay. ToxR domains (squares) can activate transcription of the reporter gene (CAT) if brought together by the GpA-derived TM domains (right). The maltose binding protein domain (ellipses) helps direct the insertion of the construct into the membrane, complements the *malE* mutation in the host cells, and serves as an epitope for quantifying the expression level of fusion protein. (B) Complementation assays for wild-type and selected mutant ToxR(GpA)MBP fusion constructs. NT326 cells (*malE* deficient) carrying various constructs were streaked on a plate with maltose as the sole carbon source and grown for three days at 37°C. All ToxR(GpA)MBP chimeras permit growth of NT326 cells on maltose, while control transformants (pccKAN) do not. (C) Normalized dimerization of the indicated TM domain variants as measured by CAT-ELISA relative to the wild-type GpA TM domain. Bars for intermediate dimerization and non-dimerizing levels are colored orange and red, respectively. Average ± s.d. of results from four independent experiments are shown. Levels of expression of selected ToxR(GpA)MBP constructs as analyzed by immunoblotting are shown at the bottom.
doi:10.1371/journal.pone.0044263.g005

SDS-PAGE analysis

Purified SN/GpA proteins were loaded onto SDS 12% polyacrylamide mini-gels. The loading buffer contained 2% (w/v) SDS, and samples were boiled for five minutes prior to electrophoresis. Gels were stained with Coomassie blue, and the percentages of monomer and dimer were estimated with a ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare). Gels with radioactive Lep-derived samples were dried at 80°C and scanned using a Fuji FLA-3000 phosphor-imager using the ImageGauge software.

ToxCAT methods

Plasmids encoding ToxR(GpA)MBP chimerae were transformed into *Escherichia coli* NT326 cells (kindly provided by D. M. Engelman) and plated onto Luria Bertani (LB) plates (with 50 μg/ml ampicillin, 25 μg/ml streptomycin); colonies were inoculated into LB medium (with 50 μg/ml ampicillin, 25 μg/ml streptomycin), and glycerol stocks were made at $A_{600}\approx0.2$ and stored at −80°C. LB cultures (with 50 μg/ml ampicillin, 25 μg/ml streptomycin) were inoculated from frozen glycerol stocks and grown at 37°C until approximately $A_{420}\approx0.6$, when culture densities were equalized by dilution into fresh culture tubes, and 6.0 A_{420} units of cells were harvested by centrifugation and washed with 0.4 ml of sonication buffer (25 mM Tris-HCl, 2 mM EDTA, pH 8.0) [41]. Cells were then resuspended in 0.6 ml of sonication buffer and lysed by probe sonication. After removing an aliquot (20 μl) for Western blot analysis, the remaining lysate was clarified by centrifugation at 13,000×g, and the supernatant was stored on ice until the spectrophotometric assay was performed. All constructs conferred the ability to grow on maltose plates to the *malE*$^-$ strain NT326, which indicates that proper membrane insertion of the ToxR(GpA)MBP fusion protein has occurred [33]. For maltose complementation assays, *E. coli* NT326 cells expressing ToxR(GpA)MBP constructs were streaked on M9 minimal media plates containing 0.4% maltose as the only carbon source, and incubated for 3 days at 37°C. All constructs showed similar expression levels of ToxR(GpA)MBP fusion protein as determined by Western blot using an anti-MBP antibody. The self-association ability of the TM domain triggers expression of a chloramphenicol transferase (*cat*) gene reporter and production of CAT protein can be quantified by a CAT-ELISA kit (Roche Diagnostics) [42]. CAT measurements and construct expression measurements were performed in at least triplicate and were normalized for the relative expression level of each construct using Western blotting [43]. All constructs showed similar expression levels of ToxR(GpA)MBP fusion proteins as determined by Western blot using an anti-MBP antibody. For Western blots samples were mixed with equal volumes of 2× SDS-PAGE sample buffer heated to 95°C for 10 min, separated on 10% (w/v) polyacrylamide mini-gels, blotted onto nitrocellulose membranes, and blocked in skim milk. ToxR(GpA)MBP chimera were detected with biotinylated anti-MBP primary antibody (NEB) and visualized with streptavidin-horseradish peroxidase conjugate and ECL reagent (GE Healthcare). Bands were quantified with an ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare).

Supporting Information

Figure S1 TM segments and flanking residues sequences. The primary sequences of the GpA TM regions used in both the SDS-PAGE and ToxCAT analyses are shown. Hydrophobic residues are boxed in yellow and flanking residues are highlighted (italic).
(EPS)

Acknowledgments

We thank D.M. Engelman (Yale University) for ToxCAT vectors and *E. coli* NT326 cells, and M.D. Oliver for preliminary results.

Author Contributions

Conceived and designed the experiments: MO IM. Performed the experiments: MB-P. Analyzed the data: MB-P CB-D MAM-R CA IM. Contributed reagents/materials/analysis tools: MB-P CB-D. Wrote the paper: MAM-R IM.

References

1. Popot JL, Engelman DM (1990) Membrane protein folding and oligomerization - The 2-stage model. Biochemistry 29: 4031–4037.
2. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433: 377–381.
3. Baeza-Delgado C, Marti-Renom MA, Mingarro I (accepted) Structure-based statistical analysis of transmembrane segments. Eur Biophys J. DOI 10.1007/s00249-012-0813-9
4. 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.
5. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450: 1026–1030.
6. Zhou FX, Merianos HJ, Brunger AT, Engelman DM (2001) Polar residues drive association of polyleucine transmembrane helices. Proc Natl Acad Sci U S A 98: 2250–2255.
7. Gratkowski H, Lear JD, DeGrado WF (2001) Polar side chains drive the association of model transmembrane peptides. Proc Natl Acad Sci U S A 98: 880–885.
8. 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.
9. Chin CN, von Heijne G (2000) Charge pair interactions in a model transmembrane helix in the ER membrane. J Mol Biol 303: 1–5.
10. DeGrado WF, Gratkowski H, Lear JD (2003) How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. Protein Sci 12: 647–665.
11. Mackenzie KR (2006) Folding and stability of alpha-helical integral membrane proteins. Chem Rev 106: 1931–1977.
12. Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM (1992) Sequence specificity in the dimerization of transmembrane α-helices. Biochemistry 31: 12719–12725.
13. Mingarro I, Whitley P, Lemmon MA, von Heijne G (1996) Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix. A rapid way to map helix-helix interactions in integral membrane proteins. Protein Sci 5: 1339–1341.
14. MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: Structure and implications. Science 276: 131–133.
15. Smith SO, Song D, Shekar S, Groesbeek M, Ziliox M, et al. (2001) Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. Biochemistry 40: 6553–6558.
16. Orzaez M, Salgado J, Gimenez-Giner A, Perez-Paya E, Mingarro I (2004) Influence of proline residues in transmembrane helix packing. J Mol Biol 335: 631–640.
17. Russ WP, Engelman DM (2000) The GxxxG motif: a framework for transmembrane helix-helix association. J Mol Biol 296: 911–919.
18. Senes A, Engel DE, DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr Opin Struct Biol 14: 465–479.
19. Cymer F, Veerappan A, Schneider D (2012) Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties. Biochimica et biophysica acta 1818: 963–973.
20. Jayasinghe S, Hristova K, White SH (2001) MPtopo: A database of membrane protein topology. Protein Sci 10: 455–458.
21. Illergard K, Kauko A, Elofsson A (2011) Why are polar residues within the membrane core evolutionary conserved? Proteins 79: 79–91.
22. Wong WC, Maurer-Stroh S, Eisenhaber F (2011) Not all transmembrane helices are born equal: Towards the extension of the sequence homology concept to membrane proteins. Biology direct 6: 57.

23. 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.
24. Duong MT, Jaszewski TM, Fleming KG, MacKenzie KR (2007) Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations. Journal of molecular biology 371: 422–434.
25. Herrmann JR, Fuchs A, Panitz JC, Eckert T, Unterreitmeier S, et al. (2010) Ionic interactions promote transmembrane helix-helix association depending on sequence context. Journal of molecular biology 396: 452–461.
26. Caputo GA, London E (2004) Position and ionization state of Asp in the core of membrane-inserted alpha helices control both the equilibrium between transmembrane and nontransmembrane helix topography and transmembrane helix positioning. Biochemistry 43: 8794–8806.
27. Tulumello DV, Deber CM (2009) SDS micelles as a membrane-mimetic environment for transmembrane segments. Biochemistry 48: 12096–12103.
28. Monne M, Nilsson I, Johansson M, Elmhed N, von Heijne G (1998) Positively and negatively charged residues have different effects on the position in the membrane of a model transmembrane helix. J Mol Biol 284: 1177–1183.
29. Garcia-Saez AJ, Mingarro I, Perez-Paya E, Salgado J (2004) Membrane-insertion fragments of Bcl-xL, Bax, and Bid. Biochemistry 43: 10930–10943.
30. Gafvelin G, Sakaguchi M, Andersson H, von Heijne G (1997) Topological rules for membrane protein assembly in eukaryotic cells. J Biol Chem 272: 6119–6127.
31. Vilar M, Sauri A, Monne M, Marcos JF, von Heijne G, et al. (2002) Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. J Biol Chem 277: 23447–23452.
32. Johansson M, Nilsson I, von Heijne G (1993) Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in *Escherichia coli* than in mammalian microsomes. Mol Gen Genet 239: 251–256.
33. Russ WP, Engelman DM (1999) TOXCAT: a mesure of transmembrane helix association in a biological membrane. Proc Natl Acad Sci USA 96: 863–868.
34. Kolmar H, Hennecke F, Gotze K, Janzer B, Vogt B, et al. (1995) Membrane insertion of the bacterial signal transduction protein ToxR and requirements of transcription activation studied by modular replacement of different protein substructures. EMBO J 14: 3895–3904.
35. Langosch D, Brosig B, Kolmar H, Fritz HJ (1996) Dimerization of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. J Mol Biol 263: 525–530.
36. Martinez-Gil L, Sauri A, Marti-Renom MA, Mingarro I (2011) Membrane protein integration into the ER. FEBS J 278: 3846–3858.
37. Zhang J, Lazaridis T (2009) Transmembrane helix association affinity can be modulated by flanking and noninterfacial residues. Biophysical journal 96: 4418–4427.
38. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26: 680–682.
39. Lemmon MA, Flanagan JM, Hunt JF, Adair BD, Bormann B-J, et al. (1992) Glycophorin A dimerization is driven by specific interactions between transmembrane α-helices. J Biol Chem 267: 7683–7689.
40. 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.
41. Sulistijo ES, Jaszewski TM, MacKenzie KR (2003) Sequence-specific dimerization of the transmembrane domain of the "BH3-only" protein BNIP3 in membranes and detergent. The Journal of biological chemistry 278: 51950–51956.
42. Vilar M, Charalampopoulos I, Kenchappa RS, Simi A, Karaca E, et al. (2009) Activation of the p75 neurotrophin receptor through conformational rearrangement of disulphide-linked receptor dimers. Neuron 62: 72–83.
43. Johnson RM, Rath A, Melnyk RA, Deber CM (2006) Lipid solvation effects contribute to the affinity of Gly-xxx-Gly motif-mediated helix-helix interactions. Biochemistry 45: 8507–8515.
44. Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. Nucleic Acids Research 40: D370–376.

Reviewing Process

Responses to the reviewers' comments

Original comments from the reviewers are shown in **bold.**

Responses to the comments are shown in plain text

Changes in the main manuscript are shown in red.

Response to the comments by the Reviewer #1.

In the paper "Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing" the authors study the effect of polar residues within transmembrane helices.

First they show that polar residues are found in known protein structures facing the lipids. Therafter they test the effect of such residues for dimerization and finds that as long as they are nog in the dimerization interface they can be accepted. Thereafter they study the insertion of TM helices that contains polar residues using the well established LEP system. The first results were also confirmed using ToxCAT. In summary this is an interesting study that

One major difference in this study from earlier studies are that the authors focus on lipid facing polar residues, while it is clear that earlier studies have shown that most polar residues within the membrane are buried.

Minor points

1. It has earlier been reported that polar residues within the membrane are relatively more common to be buried than exposed to lipid (see ref 21). However, here they authors claim "Approximately half of these ionizable residues (51%) are buried in the protein interior, but the rest are partly exposed to the lipid face." The difference might be due different normalization schemes or alternatively that residues that can "snorkel" out are included in this study. This difference should be commented on.

The indicated percentage (51%) refers to the ionizable residues found in the 20 helices mentioned above in the text. We apology for the confusion and have changed the text to clarify this fact (page 6):

"To gain more detailed insight into the structural role of these ionizable residues within the membrane core, we analyzed the environment of all these 20 helices. Approximately half of the ionizable residues (51%) found in these helices are buried in the protein interior, but the rest are partly exposed to the lipid face."

2. The authors should related all the insertion frequence of constructs in the lep study with the predicted insertion frequencies from the biological hydrophobicity scale.

We have included these data in a new Table (Table 1) and profusely discussed on page 12.

"Furthermore, the predicted insertion frequencies from the biological hydrophobicity scale [2,5] for these mutants using the ΔG Prediction Server v1.0

(http://dgpred.cbr.su.se/) are shown in Table 1. In this algorithm, the predicted insertion frequency comes from the apparent free-energy difference (ΔGapp) from insertion into ER membranes. Since very low and very high insertion efficiencies cannot be accurately measured, ΔGapp values outside the interval ±1.5 kcal/mol are only qualitative. The positive value of ΔGapp predicted that the tested-sequence is not TM. The high negative value for the GpA wild-type sequence agrees with our experimentally measured glycosylation data showing the highest insertion efficiency. A closer analysis of the output data highlighted I85D mutation as precluding TM disposition. Hence, replacing Ile85 with aspartic acid reduced ΔGapp by almost 2 kcal/mol (Table 1), which correlates with our lowest glycosylation efficiency. However, replacing Leu89 with lysine has a lower energy cost (ΔGapp close to 0) that is reflected by a higher insertion level (Fig. 4C). Finally, the double mutant I85D/L89K results in the highest predicted penalty for TM disposition, whereas experimentally we find no evidence that GpA TM segment is significantly compromised by the presence of two poplar/ionizable residues. Such phenomena points towards an intra-helical interaction between the ionizable residues and should be taken into account to improve TM prediction algorithms."

3. The authors should make it clearer that not only polar residues in the dimerization interface breaks the dimerization. (second sentence to last in conclusion).

According to the reviewer suggestion we have adapted the sentence in the conclusion section (Page 14), and to emphasize this point we prepared a new mutant (I85K/L89D) that has now been included in Fig. 3B and its dimerization degree discussed on Page 9 (first paragraph).

Page 14 "Interestingly, L89D mutant, which precludes dimer formation in the presence of SDS micelles (Fig. 3C), appears to retain some dimerization capacity (21±4%, normalized dimerization), which highlights the influence of the specific lipid environment during the assembly of TM segments [36]. Nevertheless, differences in TM segment length and flanking residues sequences (see Supplemental Fig. S1) may alter the dimerization process in the two systems, which are difficult to rationalize."

Page 9. "Furthermore, when charge pairs include L89D mutation although facing the lipids, as for I85K/L89D, we found no evidence for dimer formation (Fig. 3B)."

Response to the comments by the Reviewer #2.

In this submission, Mingarro and colleagues survey the prevalence of potentially charged side chains in a collection of hydrophobic membrane spans and then use several types of experiments to probe how the introduction of such residues influence membrane insertion and self-association (in membranes or mimetics) of the well-studied glycophorin A TMD. The reported data were acquired using methods that are well accepted in the field, and the types of conclusions that can be drawn from the results will help advance our understanding of how potentially charged residues affect the insertion and interactions among of otherwise hydrophobic TMDs in membranes. A few small changes need to be made to ensure that the data could be replicated by others.

1. For the SDS-PAGE, ToxCAT, and Lep data, the authors should provide the full amino acid sequences of the wild type GpA TMDs and flanking regions as cloned into the experimental fusion proteins for the reader, as part of each figure or ideally as a separate sequence alignment (perhaps as a supplemental figure).

We have included the TM plus flanking regions sequences used in the SDS-PAGE and ToxCAT assay as a supplemental figure as suggested by the referee, and for the Lep assay in Fig. 4A. In addition, we also commented these differences in the main text (page 13).

"Interestingly, L89D mutant, which precludes dimer formation in the presence of SDS micelles (Fig. 3C), appears to retain some dimerization capacity (21±4%, normalized dimerization), which highlights the influence of the specific lipid environment during the assembly of TM segments [36]. Nevertheless, differences in TM segment length and flanking residues sequences (see Supplemental Fig. S1) may alter the dimerization process in the two systems, which are difficult to rationalize."

2. The methods for the Lep-based experiments are not fully explained; the authors should directly cite a reference for the experimental method or provide a description. With these changes, the manuscript will meet the data quality and reporting standards of PLoS ONE.

We have explained in more detail the insertion experiments using the Lep system (pages 10-11) and added a reference in which the assay was fully described (current reference 32).

"An engineered glycosylation site placed at the C-terminal P2 domain is glycosylated efficiently upon correct insertion into the microsomal membrane (Fig. 4B), serving as a reporter to distinguish between a lumenal (glycosylated) and a cytoplasmic (unglycosylated) location. Glycosylation of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep expressed in the absence of microsomes. The efficiency of glycosylation of Lep under standard conditions is 80–90% [31,32]. The strength of the Lep system is that it provides a comparative scale for the energetic cost of inserting a broad range of model and actual TM sequences into biological membranes, closely mimicking the in vivo situation."

The conclusions drawn by the authors are largely supported by the data and are well connected to the existing literature, but the quality and value of the manuscript could be improved by addressing a few points in the Results and Conclusion (see below). Whereas each of the individual data points add to our understanding, I note in passing that had the experiments been undertaken more systematically, a better comparison of the effects in the different methods could have been made. Aspects that should be addressed in the manuscript include:

3. The authors discuss the general effects of 'ionizable residues' but the designed mutations are strongly skewed towards D and K, with only one instance of a mutation to E and two mutations to R. K is a bit more abundant than R in their TMD survey, but E is more common than D. Why the de-

emphasis of residues E and R? Can the conclusions be considered general without sampling these side chains as completely as D and K?

While we agree with the Reviewer that analyzing more mutants could potentially widened our knowledge, we note that current data do not support an expected difference between residues with the same ionization state. So, we assume that D residues will closely mimic the behavior of E residues and similarly for the positively charged residues. Sampling all the polar residues in every GpA TM position is beyond the scope of the current work. Moreover, and agreeing with the Reviewer, we trust that the present data significantly increases our understanding of the effect of polar residues in TM helix packing.

4. The authors suggest that I85D is more disruptive than L89D because "L89D mutation alters the interaction of the protein with the negatively charged detergent micelle" in a way that 85D presumably does not. Although this suggestion may be correct, the monomers of 85D and 89D in Figure 3 seem to migrate at very similar rates, and rather differently from wild type or the K mutants. Similar migration seems to indicate that the monomers associate with SDS similarly, not differently. Can this point be addressed?

We concur with the Reviewer that some gel shift could appear when non-dimerizing sequences containing polar residues are analyzed through SDS-PAGE. However, some caution should be taken in order to avoid over-interpretation of this type of information as demonstrated for example by Walkenhorst *et al.* (2009) *BBA* 1788:1321-31.

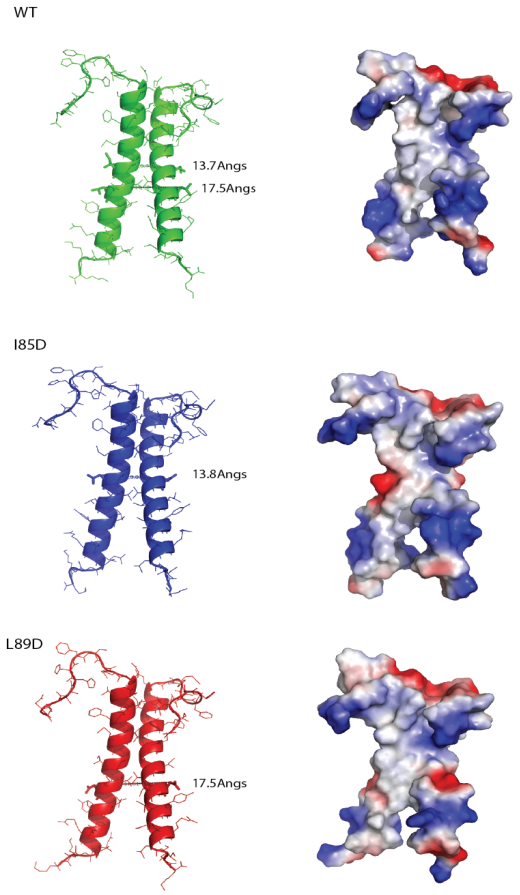
Nevertheless, we included a comment to emphasize Reviewer's point (page 8).

"Nevertheless, the comparable electrophoretic migration observed for I85D and L89D (Fig. 3C) suggests that the monomers associate with SDS quite similarly."

5. Because the GpA helices cross near residue 83, D residue carboxyls in a dimer likely approach one another more closely when at positions 85 of than when at positions 89. Have the authors modeled the mutations on the GpA structure and measured the distances? For homotypic ionizable residues there should be a strong electrostatic repulsion even if the side chains can be sterically accommodated - so if 85D brings the carboxylates closer than 89D, it would be even more surprising that 85D does not disrupt dimerization but 89D does. On the other hand, if the details of the structure place the carboxylates close together for 89D and farther apart for 85D (despite their positions in the primary sequence), then electrostatic repulsion could help explain the observed differences. This would mean that instead of 'at the interface' vs 'away from the interface', which seems to be the code for mutations that affect steric packing, the distance-dependence of electrostatic repulsion might provide a longer range effect. Even if this approach doesn't provide 'the answer', it is worth testing and accepting/rejecting given the history of the system.

We have used the MODELLER package (Sali and Blundell 1993 *JMB* 234:779-815) to model the I85D and L89D mutations using as a template structure the GpA wild-type structure (PDB code: 1afo). The MODELLER package was used with default parameters and an exhaustive minimization protocol that ensures proper conformational search of the side-chain of the mutated residue. The resulting models

for I85D and L89D (see embedded image) result in no differences in terms of distances that could explain the packing data discrepancies found in SDS. Moreover, the electrostatics surfaces calculated from the models indicate that both aspartic acid residues (red patches) are placed away from the interface. We also want to mention that only specific mutations facing the lipids allow TM-packing (see also point 3 Reviewer #1).



Given the nature of comparative protein structure approaches, which tend to reproduce the structure of the template, and the insignificant differences in the models for the mutations compared to the model of the wild-type structure, we propose not to include these results in the main text.

6. Figure 3 indicates that mutations I85K, L89K, and L89R cause the GpA TMD to dimerize MORE TIGHTLY than wild type. Do the authors have an explanation for this? (I am at a loss, especially given the argument about elecetrostatic repulsion above.) This is particularly thorny because in ToxCAT, the substitution L89K disrupts dimerization to an intermediate level. How are we to reconcile these observations, and what general (versus lipid-dependent)

conclusions can be drawn? What happens to the mutants I85K and L89R in ToxCAT?

We believe that the positively charged residues in all these mutants will gain some stabilization from the negatively charged SDS moiety, and this effect could imbalance the equilibrium towards the dimer form. Nevertheless, since the monomer bands for these mutants show some 'smearing' effect (see Fig. 3C) that somehow complicates the quantifications, we decided not to emphasize it in the main text.

7. Given the above questions, it is too bad that the ToxCAT data of Figure 5 do not include the mutant 89D (which the authors show is disruptive in SDS gels) for comparison with 85D (which is dimeric in SDS gels). Comparing the effects in SDS and ToxCAT for as many mutants as possible would be very valuable.

We have prepared L89D mutant for the ToxCAT assay and the new data are included in the current version (page 13).

"Interestingly, L89D mutant, which precludes dimer formation in the presence of SDS micelles (Fig. 3C), appears to retain some dimerization capacity (21±4%, normalized dimerization), which highlights the influence of the specific lipid environment during the assembly of TM segments [36]. Nevertheless, differences in TM segment length and flanking residues sequences (see Supplemental Fig. S1) may alter the dimerization process in the two systems, which are difficult to rationalize."

8. One technical comment: in panel C of Figure 4, the quantification of band intensisites seems to be off. In paticular, the L89K mutant seems to be quantitatively glycosylated but is reported as 79%. Has a background subtraction been performed?

Yes, the background was subtracted to perform the quantification analysis. The values shown are an average of four independent transcription/translation experiments in the presence of microsomal membranes. Please, note that mutant L89K results the largest standard deviation (±9), which raises the value till the levels shown in the image, where in fact a band corresponding to non-glycosylated molecules is still present.

Article 5:

Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co-translational Integration into and Budding Vesicle Exit from the ER Membrane

Article 5

Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co-translational Integration into and Budding Vesicle Exit from the ER Membrane

Peter U. Mayerhofer[1,4,*], Manuel Bañó-Polo[3], Ismael Mingarro[3], and Arthur E. Johnson[1,2,*]

[1]Department of Molecular and Cellular Medicine, Texas A&M Health Science Center, College Station, TX 77843, USA. [2]Departments of Chemistry and of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA. [3]Departament de Bioquimica i Biologia Molecular, Universitat de Valencia, E-46 100 Burjassot, Spain. [4]Present address: Institute of Biochemistry, Biocenter, Goethe University Frankfurt, Max-von-Laue Str. 9, 60438 Frankfurt, Germany.

The long-standing paradigm that all peroxisomal proteins are imported post-translationally into preexisting peroxisomes has been challenged by the detection of peroxisomal membrane proteins (PMPs) in the endoplasmic reticulum (ER) membrane. However, the detailed mechanism of PMP insertion into the ER is still unknown. We now show that the human PMP PEX3 inserts co-translationally into the mammalian ER via the Sec61 translocon. Photocrosslinking and fluorescence spectroscopy studies demonstrate that the N-terminal transmembrane segment of ribosome-bound PEX3 is recognized by the signal recognition particle (SRP). SRP mediates the targeting of PEX3-containing ribosome•nascent chain complexes to the translocon, where an ordered pathway integrates nascent PEX3 into the membrane adjacent to translocon proteins Sec61α and TRAM. PEX3 insertion into ER membranes is physiologically relevant because PEX3 exits the ER via budding vesicles in an ATP-dependent process. These sequential stages of human PMP passage through the mammalian ER therefore delineate early steps in peroxisome biogenesis.

INTRODUCTION

The significance of peroxisomes in cellular metabolism is illustrated by the existence of severe inherited human diseases that result from the failure of peroxisomal biogenesis (Lazarow and Moser, 1995). More than 30 proteins (termed peroxins) are involved in peroxisomal biogenesis across species (reviewed in Fujiki et al., 2006; Ma et al., 2011; Rucktaschel et al., 2011), but only three are key players in early peroxisomal membrane assembly. PEX19 is a soluble protein that acts as receptor and chaperone for newly synthesized PMPs in the cytosol (Jones et al., 2004). PEX16 is a membrane protein, but homologues are absent in most yeast species (Fujiki et al., 2006). The PEX3 PMP is highly conserved among species and has been proposed to be the docking factor for cytosolic PEX19•cargoPMP complexes (Fang et al., 2004; Matsuzaki and Fujiki, 2008).

Peroxisomes have long been considered to be autonomous organelles that arise exclusively by growth and division of pre-existing peroxisomes (Fujiki et al., 1984; Lazarow and Fujiki, 1985). However, convincing evidence has recently shown that at least a subpopulation of PMPs in yeast (Baerends et al., 1996; Elgersma et al., 1997; Hoepfner et al., 2005; Thoms et al., 2012; Titorenko et al., 1997; van der Zand et al., 2010), plant (Mullen et al., 1999), and vertebrate cells (Geuze et al., 2003; Kim et al., 2006; Toro et al., 2009; Yonekawa et al., 2011) are targeted first to the ER prior to being transported to the peroxisomes via an ER-derived vesicle carrier (Agrawal et al., 2011; Lam et al., 2010; van der Zand et al., 2012). As a prerequisite to understanding these early steps in peroxisomal biogenesis, it is essential to ascertain how peroxins are inserted into the ER membrane. In yeast, two pathways have been identified as being involved: a small group of tail-anchored PMPs are most likely post-translationally inserted via the GET3 pathway (Schuldiner et al., 2008; van der Zand et al., 2010), while others appear to be inserted through the yeast Sec61p translocon (Thoms et al., 2012; van der Zand et al., 2010) that serves as the primary ER entry point for integral membrane and secretory proteins. Depending on its exact protein composition, the yeast Sec61p complex promotes co- and post-translational targeting of proteins (Rapoport, 2007; Shao and Hegde, 2011), and both mechanisms might be involved in the translocon-mediated insertion of PMPs into the yeast ER (Thoms et al., 2012). However, previous experiments (Thoms et al., 2012; van der Zand et al., 2010) did not reveal any mechanistic details about how the yeast Sec61p translocon facilitates PMP insertion into the ER bilayer. We have therefore focused on identifying the mechanisms involved in human PEX3 targeting to and insertion into the mammalian ER membrane.

RESULTS AND DISCUSSION

As soon as a cleavable signal sequence or an uncleaved signal-anchor sequence of an integral membrane protein emerges from the ribosome, it is recognized and bound by the SRP (reviewed in Rapoport, 2007; Shao and Hegde, 2011). This interaction transiently arrests protein synthesis until the SRP interacts with its ER-resident receptor to target the ribosome•nascent chain complex (RNC) to a

translocon in the ER membrane. Two hydrophobic regions, HR1 and HR2 (Fig. 1A) have been identified in human PEX3 (Kammerer et al., 1998). Since HR1 emerges first from the ribosomal tunnel during ribosomal synthesis (Fig. 1B), HR1 interactions were examined after leaving the tunnel by using environmentally sensitive probes. A photoreactive crosslinking probe (5-azido-2-nitrobenzoyl, ANB) or a fluorescent dye (7-nitrobenz-2- oxa-1,3-diazole, NBD) was positioned in the middle of HR1 by in vitro translation of a PEX3 mRNA in which codon 25 was replaced by an amber stop codon (PEX3G25amb, see also Fig. 3A). Addition of amber suppressor aminoacyl-tRNA analogs εANB-Lys- tRNAamb or εNBD-Lys-tRNAamb (Crowley et al., 1993; Flanagan et al., 2003; McCormick et al., 2003) to the translation then allowed selective labelling of HR1 with the probe. When a truncated PEX3 mRNA transcript lacking a final stop codon was translated, all nascent chains in the resulting RNC sample had the same length and remained attached to the ribosome as peptidyl-tRNA because normal termination was prevented. By varying the length of truncated mRNA added to translations, RNCs with different nascent chain lengths provided a series of static snapshots of sequential stages in PMP membrane targeting and integration. Nascent chains are designated P(x)-PEX3(n) to represent PEX3 nascent chains with a length of n residues and a probe P at residue x. Since the ribosomal exit tunnel encloses roughly 35 residues of an emerging nascent chain, HR1 will be fully exposed to the cytosol in RNCs containing PEX3(93) (Fig. 1B).

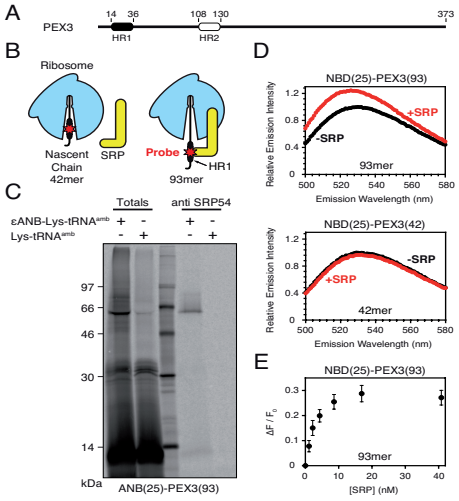


Figure 1. HR1 of PEX3 binds to SRP (A) Schematic representation of full-length PEX3. Two predicted hydrophobic α-helical regions are indicated by black (HR1) and white (HR2) boxes. (B) A probe (either a photo-reactive crosslinker or a fluorescent dye) is incorporated into a specific site within HR1 of ribosome-tethered nascent PEX3. The HR1 of short nascent chains (e.g. 42mer) is located within the ribosomal exit tunnel, whereas longer chains (e.g. 93mer) expose HR1 to the cytosol and hence to the signal recognition particle (SRP). (C) Photocrosslinking of PEX3 to SRP. [35S]Met-labelled PEX3-RNCs with a nascent chain length of 93 residues and a single photoreactive ANB probe in HR1 were photolyzed and then analyzed by SDS-PAGE and phosphorimaging either directly (Totals, 1/20 aliquot) or after immunoprecipitation with antibodies directed against SRP54. (D) Fluorescence-detected SRP binding to PEX3. Truncated PEX3G25amb mRNA was translated in the presence of εNBD-Lys-tRNAamb. Emission scans (lex = 468 nm) of purified NBD(25)-PEX3-RNCs were performed in buffer A before (-SRP) and immediately after the addition of purified canine SRP (+SRP). (E) NBD(25)-PEX3(93)-RNCs were titrated with the indicated total concentrations of SRP. The observed change in emission intensity (lex = 468 nm; lem = 528 nm) is ΔF, and the initial fluorescence intensity of the sample without SRP is designated F0. The averages of at least three independent experiments are shown, with error bars indicating the S.D.

Photoreactive ANB was introduced into HR1 by translating a truncated PEX3G25amb mRNA in the presence of εANB-Lys-tRNAamb; the control sample received Lys-tRNAamb. The resulting ANB(25)-PEX3(93) and PEX3(93) RNCs were photolyzed, and a prominent photoadduct was formed only in the sample with ANB (Fig.1C).Photoadducts were then analyzed by immunoprecipitation using antibodies specific for SRP54, the signal sequence-binding component of SRP (Krieg et al., 1986; Kurzchalia et al., 1986). Since [35S]Met-labelled ANB(25)-PEX3(93) chains reacted covalently with SRP54 (Fig. 1C), the photoreactive ANB in HR1 was adjacent to SRP54. Thus, HR1 was recognized and bound by SRP as it emerged from the ribosome.

The association of SRP with PEX3-containing RNCs was also detected using a NBD fluorescent probe in HR1. NBD was chosen because its emission properties change dramatically upon moving from an aqueous to a hydrophobic environment (Crowley et al., 1993), and we previously showed that NBD was a sensitive spectral sensor of SRP association with a RNC signal sequence (Flanagan et al., 2003). NBD was introduced at position 25 of HR1 by translating truncated PEX3G25amb mRNA in the presence of εNBD-Lys-tRNAamb. When canine SRP was added to purified NBD(25)- PEX3(93)-RNCs, a significant increase in NBD emission intensity was observed (Fig. 1D, top). In contrast, no increase in emission intensity was detected when only buffer was added to NBD(25)-PEX3(93) RNCs (Fig. S1) or when SRP was incubated with NBD(25)-PEX3(42) RNCs (Fig. 1D, bottom) with HR1 still inside the ribosomal exit tunnel (Fig. 1B). SRP binding to NBD(25)-PEX3(93)-RNCs was saturable, as shown by the dependence of sample emission intensity on the concentration of SRP (Fig. 1E).

The HR1 interaction with SRP indicates that the nonpolar HR1 acts as a signal sequence. Does HR1 also function as a transmembrane segment (TMS) to anchor PEX3 in the membrane? PEX3 HR segments were engineered into the Escherichia coli inner membrane protein leader peptidase (Lep; Hessa et al., 2005; Fig. 2A), and the glycosylation pattern revealed that isolated HR1, but not HR2, was efficiently integrated into the ER membrane (Fig. 2B). Furthermore, carbonate extraction of PEX3 and a derivative lacking HR1 (Fig. 2C) showed that HR1 is necessary (Fig. 2D) and sufficient (Fig. S2) for stable insertion of PEX3 into the ER bilayer. The large C-terminal domain of ER-inserted PEX3 was therefore exposed to the cytosol (Fig. 2E), a topology previously described for peroxisomal-localized PEX3 (Kammerer et al., 1998; Soukupova et al.,

1999). Since ER membrane-integrated and non-inserted PEX3 had identical molecular masses (Fig. 2D, E), the N-terminal hydrophobic HR1 of PEX3 acts as a non-cleavable signal-anchor TMS that is recognized by SRP. These data therefore provide the first direct evidence that the TMS of a peroxisomal integral membrane protein can be recognized and bound by the SRP when it emerges form the ribosomal tunnel.

Following SRP•RNC docking at the ER membrane, the Sec61 translocon mediates both the transport of soluble proteins into the ER lumen and the insertion of integral membrane proteins into the ER bilayer (Shao and Hegde, 2011). The mammalian Sec61 translocon is composed of four core proteins, Sec61α,β,γ and the translocating chain-associating membrane protein (TRAM) (Johnson and van Waes, 1999). To examine SRP dependence of PEX3 targeting to the translocon, PEX3-RNCs were translated in a wheat germ extract that has such a low endogenous content of SRP that RNC targeting to column-washed canine rough microsomes (CRM) is dependent on added canine SRP (Krieg et al., 1986; Tamborero et al., 2011). ANB(25)-PEX3(93) RNCs were translated in the presence of CRM, εANB-Lys-tRNAamb, and either the presence or absence of SRP. After photolysis and immunoprecipitation using antibodies specific for Sec61α, covalent photoadducts between Sec61α and PEX3 nascent chains were observed only in the presence of SRP (Fig. 3B). No photoadducts were observed in the absence of the photoreactive probe (data not shown). Thus, SRP is required to target nascent PEX3 to the translocon.

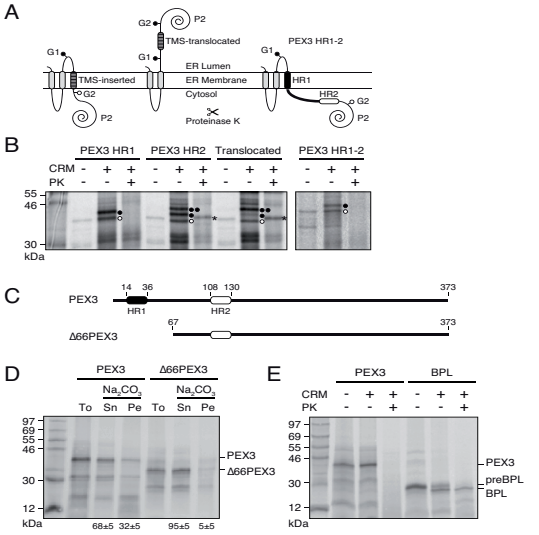


Figure 2. HR1 is responsible for ER membrane insertion of PEX3
(A) Schematic diagram of E. coli leader peptidase (Lep) constructs used to test TMS insertion into the ER bilayer. The putative TMS (shaded) is engineered into the P2 domain of Lep flanked by two artificial glycosylation acceptor sites (G1 and G2). Membrane integration of the TMS prevents enzymatic glycosylation of G2 (m) on the lumenal side of the membrane, whereas both sites are glycosylated (l) when a TMS does not insert into the membrane. (B) Insertion of PEX3 HR1 (residues 14-36), HR2 (residues 108-130), or HR1-HR2 (residues 14-130) fragments into ER membranes. PEX3-HR-Lep chimeras or a control shown previously to be translocated (construct no. 67; Martinez-Gil et al., 2010) were transcribed and translated in rabbit reticulocyte lysate in either the presence or absence of microsomal membranes (CRM). [35S]Met-labelled translation products were analyzed directly or treated with Proteinase K (PK). Unglycosylated proteins are indicated by open circles; mono- and double-glycosylated proteins are indicated by one and two filled circles, respectively. Stars depict protease- protected fragments. (C) Schematic representation of full-length PEX3 and a truncated variant lacking the N-terminal 66 residues (D66PEX3). (D) Full-length PEX3 is anchored in the ER bilayer. PEX3 variants were translated in rabbit reticulocyte lysate in the presence of CRMs. Translation products were subjected to sodium carbonate extraction at pH 11.5 and separated by centrifugation. The supernatant (Sn), the membrane pellet (Pe), and an untreated aliquot (To) are shown. Numbers indicate the average amount of PEX3 or D66PEX3 in the supernatant and membrane pellet fractions, respectively. The averages ± S.D. of at least 3 independent experiments are shown. (E) Orientation of ER- inserted PEX3. Full-length PEX3 or secreted bovine prolactin (BPL) were translated as above in either the absence or presence of CRM. Translation products were analyzed directly or treated with PK. pre, BPL with an uncleaved signal sequence.

To further characterize HR1 interactions at the translocon, we used a high- resolution photocrosslinking approach. Parallel samples of ANB(23)-PEX3(79), ANB(24)-PEX3(79), and ANB(25)-PEX3(79) integration intermediates were generated and photolyzed, and the extent of photocrosslinking to translocon proteins was determined by immunoprecipitation with antibodies specific for Sec61α and TRAM. The ANBs incorporated at three sequential residues within HR1 project from three different sides of the TMS α-helix (Fig. 3A). If HR1 is randomly oriented when it is proximal to membrane integration) of the molecule results in an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of Lep expressed in the absence of microsomes, and of around 5 kDa in the case of double glycosylation (i.e., membrane translocation of the H-segment). Proteinase K (PK) added to MM vesicles will digest the cytoplasmic exposed, non-glycosylated form of the P2 domain (Fig. 1A, left) and produces a protected, doubly-glycosylated P2 fragment when the P2 domain is located in the lumen of the MMs (Fig. 1A, right).

Sec61α, then all three probes should react equally with Sec61α and/or TRAM. However, if an asymmetric photocrosslinking pattern is observed, then HR1 must be held in a fixed orientation adjacent to Sec61α and/or TRAM (McCormick et al., 2003). Since only probes at residue 25 of PEX3 photocrosslinked to Sec61α (Fig. 3C), probes at both positions 24 and 25 photocrosslinked to TRAM (Fig. 3D), and probes at residue 23 photocrosslinked to neither translocon protein, the asymmetry of photocrosslinking reveals that HR1 is bound to a specific site within the translocon.

HR1 proximity to translocon proteins was then examined as a function of nascent chain length. Since an εANB-Lys at PEX3 residue 25 photocrosslinked to both Sec61α and TRAM, ANB(25)-PEX3 RNCs with increasing nascent chain lengths were prepared in parallel, photolyzed, and analyzed by immunoprecipitation. When nascent chain

length increased beyond 93 residues, HR1 was no longer adjacent to Sec61α (Fig. 3E, F). TRAM-containing photoadducts were observed with nascent chain lengths of 93 and 122, 148 to a lesser extent (Fig. 3F), and not at all for nascent chains 192 or more residues (Fig. S3). Since HR1 was adjacent to TRAM, but not to Sec61α, at 122 residues, HR1 was retained next to TRAM longer than to Sec61α, consistent with earlier data showing a TMS passing sequentially from Sec61α to TRAM during integration at the translocon (Do et al., 1996; Sauri et al., 2007). PEX3 therefore inserts co- translationally into the ER membrane via the standard SRP-dependent and translocon- mediated multistep pathway.

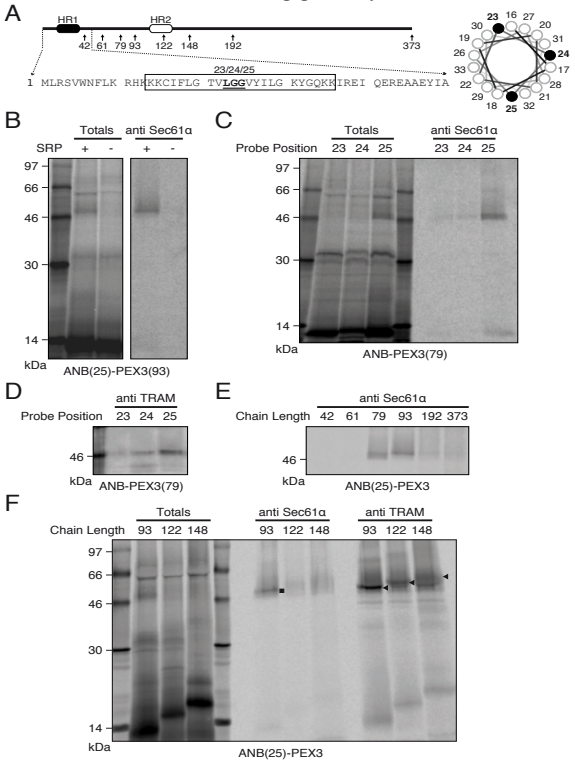


Figure 3. Photocrosslinking of nascent PEX3 to the translocon proteins Sec61α and TRAM (A) Schematic diagram and N-terminal sequence of human PEX3. Arrows indicate PEX3 nascent chains of different lengths used in this study. An amber stop codon was substituted for the codon at position L23, G24, or G25 (underlined) to position the photoreactive ANB at a single nascent chain location within HR1 (boxed). When the probe was incorporated at 23, 24, or 25 in different samples, the probe projected from different sides of the TMS α-helix surface as shown in the helical wheel projection. (B) Photocrosslinking to Sec61α is SRP-dependent. [35S]Met-labelled ANB(25)-PEX3(93) nascent chains were prepared in wheat germ extract supplemented with canine CRMs in either the absence or presence of canine SRP. After photolysis, photoadducts were analyzed directly (Totals, 1/20 aliquot) or after immunoprecipitation with antibodies specific for Sec61α. (C, D, E, F) Photocrosslinking to Sec61α and TRAM. [35S]Met- labelled ANB(23)-PEX3, ANB(24)-PEX3, or ANB(25)-PEX3 integration intermediates were translated in lengths of 79 (C, D) or 42, 61, 79, 93, 122, 148, 192, and 373 (full- length) residues (E, F) in the presence of CRMs and SRP. After photolysis, photoadducts were analyzed either directly (Totals, 1/20 aliquot) or after immunoprecipitation with antibodies directed against Sec61α or TRAM, respectively. Uncropped images of D and E are shown in Fig. S4.

Does every human PEX3 PMP insert into the ER membrane via the SRP- and translocon-mediated pathway? Given the sub-stoichiometric number of SRPs relative to ribosomes (in yeast, there are 1-2 SRPs for every 100 ribosomes (Raue et al., 2007)), it is certainly possible that some PEX3 molecules may escape recognition by SRP and be inserted post-translationally into peroxisomes (Matsuzaki and Fujiki, 2008) or the ER (Kim et al., 2006; Toro et al., 2009). In such cases, the initial co-translationally inserted PEX3 may then serve as docking factor for PEX19•cargoPMP complexes (Fang et al., 2004; Matsuzaki and Fujiki, 2008) in the ER and thereby concentrate PMPs or PMP sub- complexes (van der Zand et al., 2012) in a spatially defined area of the ER.

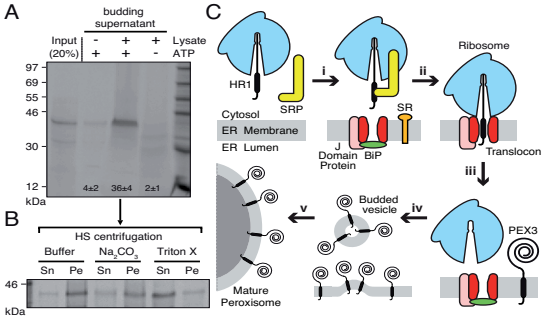


Figure 4. Cell-free vesicle budding of PEX3 (A) Full-length PEX3 was transcribed/translated in vitro in rabbit reticulocyte lysate in the presence of CRMs. Washed donor membranes were incubated at 30°C in the presence of either buffer A (– lysate) or reticulocyte lysate. Samples were either substituted with an ATP-regenerating system (+ATP) or treated with apyrase (-ATP). After the budding reaction, donor membranes were removed by sedimentation, and the supernatant fraction and a 20% aliquot (Input) of the starting microsomes were analyzed. Numbers indicate the average amount of budded PEX3 ± S.D. for at least 3 independent experiments. (B) The supernatant of an ATP- and lysate-containing budding reaction was subjected to high-speed (HS) centrifugation, and the pellet was resuspended in buffer A with or without 1% (v/v) Triton X-100 or subjected to 0.1 M sodium carbonate extraction at pH 11.5. After a second centrifugation step, the protein contents of the supernatant (Sn) and pellet (Pe) fractions were analyzed. (C) Model of human PEX3 passage through the ER. During ribosomal translation of PEX3, HR1 is recognized and bound by SRP (i). After SRP-dependent targeting of the RNC to the ER membrane (ii) via the SRP receptor (SR), PEX3 is co-translationally integrated into the mammalian ER at the Sec61 translocon and its associated proteins (J Domain Protein, BiP; Lin et al., 2011) (iii). Following integration into the ER membrane, PEX3 is selectively packed into budding vesicles in an ATP- and cytosol-dependent process (iv). PEX3-containing budded vesicles then either fuse with pre-existing peroxisomes or initiate peroxisomal de novo synthesis (v). In addition to the co-translational ER insertion, some PEX3 molecules may escape recognition by SRP and be inserted post-translationally into the ER (Kim et al., 2006; Toro et al., 2009) or peroxisomes (Matsuzaki and Fujiki, 2008). In mature peroxisomes, PEX3 serves as the membrane receptor for the post-translational insertion of cytosolic PEX19•cargoPMP complexes (Fang et al., 2004).

Is PEX3 integration into the ER membrane a precursor to PEX3 transport to the peroxisome? If so, one would predict that PEX3 is segregated into specific regions of the ER membrane for budding and transport to the peroxisome

(Yonekawa et al., 2011). A cell-free vesicle budding assay recently established in yeast (Agrawal et al., 2011; Lam et al., 2010) shows that PMP-containing carrier vesicles are released from the ER in a cytosol- and ATP-dependent process. To determine whether human PEX3 is packed into vesicles that bud from mammalian ER membranes, full-length PEX3 was translated in vitro in the presence of canine ER microsomes. Following translation, membranes were collected and washed extensively to remove any peripherally attached PEX3. These microsomes were then used as donor membranes to study the ER exit of PEX3 in the presence of rabbit reticulocyte lysate, ATP, and an ATP regenerating system. After the budding reaction, the larger and more dense donor microsomal membranes were removed by medium-speed centrifugation. PEX3 was then detected in the supernatant fraction of samples containing reticulocyte lysate and ATP, but not in the supernatant of samples lacking either cytosol or ATP (Fig. 4A). Budded PEX3 could be collected by high-speed centrifugation, was resistant to carbonate extraction, and was solubilized in detergent (Fig. 4B), thereby indicating that PEX3 was localized in a membrane of small vesicles. Since 36 ± 4% of the total integrated PEX3 was recovered in the supernatant in the presence of cytosol and ATP (Fig. 4A), PEX3 was apparently selected and preferentially transferred to the small ER-derived vesicles. These data therefore provide the first direct evidence that human PMPs are actively and selectively extracted from mammalian ER membranes in a cytosol-dependent and ATP-consuming vesicle budding reaction. Moreover, these data are consistent with small ER-derived vesicles playing a role in PEX3 trafficking to peroxisomes.

The combined data presented here establish that human PEX3 is targeted to a mammalian ER membrane, integrated co-translationally at a mammalian translocon, and then selectively packaged and extracted from the ER membrane via an energy- and cytsosol-dependent budding reaction. By experimentally characterizing the entire pathway required for PEX3 passage through the ER membrane (Fig. 4C), from recruitment and entry to exit and discharge, the transient role of the ER in mammalian peroxisomal biogenesis is now defined from entry to exit. Since it is widely accepted that certain integral PMPs, including PEX3 (Matsuzaki and Fujiki, 2008), can insert post-translationally into peroxisomes (Fang et al., 2004; Jones et al., 2004), the next goal is to determine what fraction of PEX3 molecules are inserted directly into pre-existing peroxisomes instead of transiting through the ER.

EXPERIMENTAL PROCEDURES

Photocrosslinking and Immunoprecipitation

In vitro translations of truncated mRNAs were performed in wheat germ cell-free extract in the presence of canine SRP, CRM, [35S]Met, [14C]Lys-tRNAamb / εANB-[14C]Lys-tRNAamb as indicated, and other components as described (Crowley et al., 1993). Samples were photolyzed and then analyzed by SDS-PAGE and phosphorimaging either directly or after immunoprecipitation using antibodies specific for SRP54, Sec61α, or TRAM, respectively.

Fluorescence Spectroscopy

In vitro translations of truncated mRNAs were performed in wheat germ cell-free extract in the presence of εNBD-[14C]Lys-tRNAamb/[14C]Lys-tRNAamb and other components as described (Crowley et al., 1993). After translation, RNCs were purified by gel filtration. Steady-state fluorescence measurements were performed at 4°C as described previously (Flanagan et al., 2003).

Budding Assay

Full-length PEX3 mRNA was translated in rabbit reticulocyte lysate in the presence of CRMs. After translation, membranes were washed with 2.5 M urea to remove peripherally bound PEX3 molecules. Donor membranes were incubated with either rabbit reticulocyte lysate or an equivalent amount of buffer A. Budding reactions also contained either an energy generating system or apyrase. After the budding reaction, donor membranes were removed by medium-speed centrifugation (20,000 g, 10 min, 4°C). In certain cases, the supernatant of a budding reaction was further subjected to high-speed centrifugation (Beckman TLA100 rotor; 55,000 rpm; 30 min; 4°C). The resulting membrane pellet was analyzed by carbonate extraction or solubilized with 1% (v/v) Triton X-100.

ACKNOWLEDGMENTS

We are grateful to Y. Miao and Y. Shao for technical assistance, and A. Hinz and R. Draheim for helpful discussions. This work was supported by NIH grant GM26494 (A.E.J.), by the Robert A. Welch Foundation (A.E.J., Chair grant BE-0017), and by BFU2012-39482 from the Spanish Ministerio de Economia y Competitividad (I.M.). M.B.- P. was recipient of FPU predoctoral fellowship. P.U.M., M.B.P., and I.M. declare that they have no competing financial interests. A.E.J. is a founder of tRNA Probes, LLC.

REFERENCES

Agrawal, G., Joshi, S., and Subramani, S. (2011). Cell-free sorting of peroxisomal membrane proteins from the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA 108, 9113-9118.

Baerends, R.J., Rasmussen, S.W., Hilbrands, R.E., van der Heide, M., Faber, K.N., Reuvekamp, P.T., Kiel, J.A., Cregg, J.M., van der Klei, I.J., and Veenhuis, M. (1996). The Hansenula polymorpha PER9 gene encodes a peroxisomal membrane protein essential for peroxisome assembly and integrity. J. Biol. Chem. 271, 8887-8894.

Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. Cell 73, 1101-1115.

Do, H., Falcone, D., Lin, J., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell 85, 369-378.

Elgersma, Y., Kwast, L., van den Berg, M., Snyder, W.B., Distel, B., Subramani, S., and Tabak, H.F. (1997). Overexpression of Pex15p, a phosphorylated peroxisomal integral membrane protein required for peroxisome assembly in S.cerevisiae, causes proliferation of the endoplasmic reticulum membrane. EMBO J. 16, 7326-7341.

Fang, Y., Morrell, J.C., Jones, J.M., and Gould, S.J. (2004). PEX3 functions as a PEX19 docking factor in the import of class I peroxisomal membrane proteins. J. Cell Biol. 164, 863-875.

Flanagan, J.J., Chen, J.C., Miao, Y., Shao, Y., Lin, J., Bock, P.E., and Johnson, A.E. (2003). Signal recognition particle binds to ribosome-bound signal sequences with fluorescence-detected subnanomolar affinity that does not diminish as the nascent chain lengthens. J. Biol. Chem. 278, 18628-18637.

Fujiki, Y., Matsuzono, Y., Matsuzaki, T., and Fransen, M. (2006). Import of peroxisomal membrane proteins: the interplay of Pex3p- and Pex19p-mediated interactions. Biochim. Biophys. Acta 1763, 1639-1646.

Fujiki, Y., Rachubinski, R.A., and Lazarow, P.B. (1984). Synthesis of a major integral membrane polypeptide of rat liver peroxisomes on free polysomes. Proc. Natl. Acad. Sci. USA 81, 7127-7131.

Geuze, H.J., Murk, J.L., Stroobants, A.K., Griffith, J.M., Kleijmeer, M.J., Koster, A.J., Verkleij, A.J., Distel, B., and Tabak, H.F. (2003). Involvement of the endoplasmic reticulum in peroxisome formation. Mol. Biol. Cell 14, 2900-2907.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and von Heijne, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433, 377-381.

Hoepfner, D., Schildknegt, D., Braakman, I., Philippsen, P., and Tabak, H.F. (2005). Contribution of the endoplasmic reticulum to peroxisome formation. Cell 122, 85-95.

Johnson, A.E., and van Waes, M.A. (1999). The translocon: a dynamic gateway at the ER membrane. Annu. Rev. Cell Dev. Biol. 15, 799-842.

Jones, J.M., Morrell, J.C., and Gould, S.J. (2004). PEX19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. J. Cell Biol. 164, 57-67.

Kammerer, S., Holzinger, A., Welsch, U., and Roscher, A.A. (1998). Cloning and characterization of the gene encoding the human peroxisomal assembly protein Pex3p. FEBS Lett. 429, 53-60.

Kim, P.K., Mullen, R.T., Schumann, U., and Lippincott-Schwartz, J. (2006). The origin and maintenance of mammalian peroxisomes involves a de novo PEX16-dependent pathway from the ER. J. Cell Biol. 173, 521-532.

Krieg, U.C., Walter, P., and Johnson, A.E. (1986). Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. Proc. Natl.. Acad. Sci. USA 83, 8604-8608.

Kurzchalia, T.V., Wiedmann, M., Girshovich, A.S., Bochkareva, E.S., Bielka, H., and Rapoport, T.A. (1986). The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. Nature 320, 634-636.

Lam, S.K., Yoda, N., and Schekman, R. (2010). A vesicle carrier that mediates peroxisome protein traffic from the endoplasmic reticulum. Proc. Natl. Acad. Sci. USA 107, 21523-21528. Lazarow, P.B., and Fujiki, Y. (1985). Biogenesis of peroxisomes. Annu. Rev. Cell Biol. 1, 489-530.

Lazarow, P.B., and Moser, H.W. (1995). Disorders of peroxisome biogenesis. In The Metabolic and Molecular Bases of Inherited Disease, C.R. Scriver, A.L. Beaudet, W.S. Sly, and D. Valle, eds. (New York: McGraw-Hill), pp. 2287-2324.

Lin, P.J., Jongsma, C.G., Liao, S., and Johnson, A.E. (2011). Transmembrane segments of nascent polytopic membrane proteins control cytosol/ER targeting during membrane integration. J. Cell Biol. 195, 41-54.

Ma, C., Agrawal, G., and Subramani, S. (2011). Peroxisome assembly: matrix and membrane protein biogenesis. J. Cell Biol. 193, 7-16.

Martinez-Gil, L., Johnson, A.E., and Mingarro, I. (2010). Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. J. Virol. 84, 5520-5527.

Matsuzaki, T., and Fujiki, Y. (2008). The peroxisomal membrane protein import receptor Pex3p is directly transported to peroxisomes by a novel Pex19p- and Pex16p-dependent pathway. J. Cell Biol. 183, 1275-1286.

McCormick, P.J., Miao, Y., Shao, Y., Lin, J., and Johnson, A.E. (2003). Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. Mol. Cell 12, 329-341.

Mullen, R.T., Lisenbee, C.S., Miernyk, J.A., and Trelease, R.N. (1999). Peroxisomal membrane ascorbate peroxidase is sorted to a membranous network that resembles a subdomain of the endoplasmic reticulum. Plant Cell 11, 2167-2185.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature 450, 663-669.

Raue, U., Oellerer, S., and Rospert, S. (2007). Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. J. Biol. Chem. 282, 7809-7816.

Rucktaschel, R., Girzalsky, W., and Erdmann, R. (2011). Protein import machineries of peroxisomes. Biochim. Biophys. Acta 1808, 892-900.

Sauri, A., McCormick, P.J., Johnson, A.E., and 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.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell 134, 634-645.

Shao, S., and Hegde, R.S. (2011). Membrane protein insertion at the endoplasmic reticulum. Annu. Rev. Cell Dev. Biol. 27, 25-56. Soukupova, M., Sprenger, C., Gorgas, K., Kunau, W.H., and Dodt, G. (1999). Identification and characterization of the human peroxin PEX3. Eur. J. Cell Biol. 78, 357- 374.

Tamborero, S., Vilar, M., Martinez-Gil, L., Johnson, A.E., and Mingarro, I. (2011). Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). J. Mol. Biol. 406, 571-582.

Thoms, S., Harms, I., Kalies, K.U., and Gartner, J. (2012). Peroxisome Formation Requires the Endoplasmic Reticulum Channel Protein Sec61. Traffic 13, 599-609.

Titorenko, V.I., Ogrydziak, D.M., and Rachubinski, R.A. (1997). Four distinct secretory pathways serve protein secretion, cell surface growth, and peroxisome biogenesis in the yeast Yarrowia lipolytica. Mol. Cell. Biol. 17, 5210-5226.

Toro, A.A., Araya, C.A., Cordova, G.J., Arredondo, C.A., Cardenas, H.G., Moreno, R.E., Venegas, A., Koenig, C.S., Cancino, J., Gonzalez, A., and Santos, M.J. (2009). Pex3p- dependent peroxisomal biogenesis initiates in the endoplasmic reticulum of human fibroblasts. J. Cell. Biochem. 107, 1083-1096.

van der Zand, A., Braakman, I., and Tabak, H.F. (2010). Peroxisomal membrane proteins insert into the endoplasmic reticulum. Mol. Biol. Cell 21, 2057-2065.

van der Zand, A., Gent, J., Braakman, I., and Tabak, H.F. (2012). Biochemically distinct vesicles from the endoplasmic reticulum fuse to form peroxisomes. Cell 149, 397-409.

Yonekawa, S., Furuno, A., Baba, T., Fujiki, Y., Ogasawara, Y., Yamamoto, A., Tagaya, M., and Tani, K. (2011). Sec16B is involved in the endoplasmic reticulum export of the peroxisomal membrane biogenesis factor peroxin 16 (Pex16) in mammalian cells. Proc. Natl. Acad. Sci. USA 108, 12746-12751.

Supplemental Text and Figures

Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co-translational Integration into and Budding Vesicle Exit from the ER Membrane

Peter U. Mayerhofer, Manuel Bañó-Polo, Ismael Mingarro, and Arthur E. Johnson

INVENTORY

SUPPLEMENTAL FIGURES

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

SUPPLEMENTAL REFERENCES

SUPPLEMENTAL FIGURES

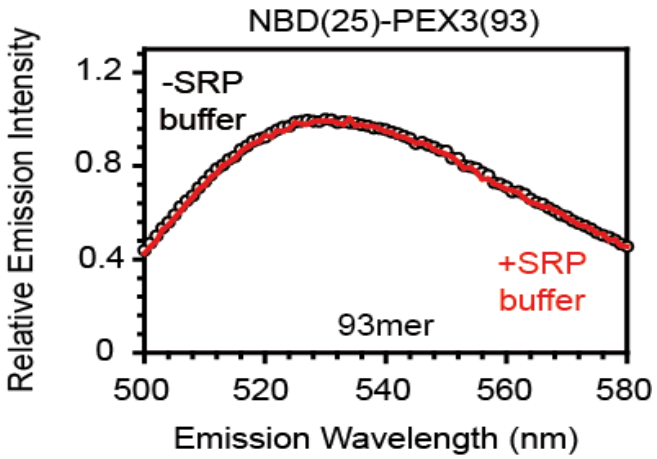


Figure S1 (related to main Figure 1). SRP storage buffer does not alter the emission intensity of fluorescence-labeled PEX3

Truncated PEX3^{G25amb} mRNA was translated in wheat germ extract in the presence of εNBD-Lys-tRNA^{amb}. Emission scans (λ_{ex} = 468 nm) of purified NBD(25)-PEX3(93)-RNCs were performed in buffer A before (-SRP buffer) and immediately after the addition of SRP storage buffer (+SRP buffer, equal volume as in Figure 1D).

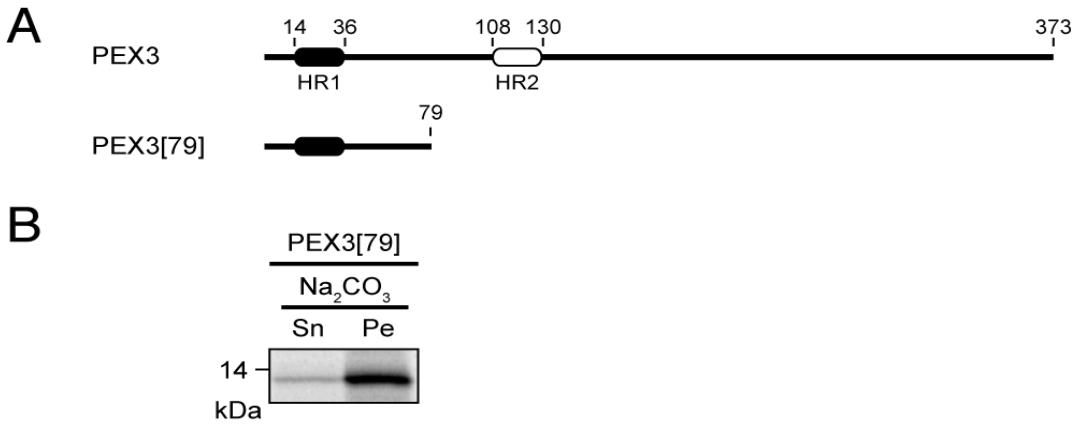


Figure S2 (related to main Figure 2). HR1 of PEX3 is stably anchored in the ER bilayer

(A) Schematic representation of full-length PEX3 and a C-terminally truncated PEX3 variant of 79 residues length (PEX[79]). Two predicted hydrophobic α-helical regions (HR) are indicated by black (HR1) and white (HR2) boxes. (B) PEX[79] was translated in rabbit reticulocyte lysate in the presence of CRMs. [^{35}S]Met-labelled translation products were subjected to sodium carbonate extraction at pH 11.5. After centrifugation (100,000 x g; 20 min), the supernatant (Sn), and the membrane pellet (Pe) were analyzed by SDS-PAGE and visualized by phosphorimaging.

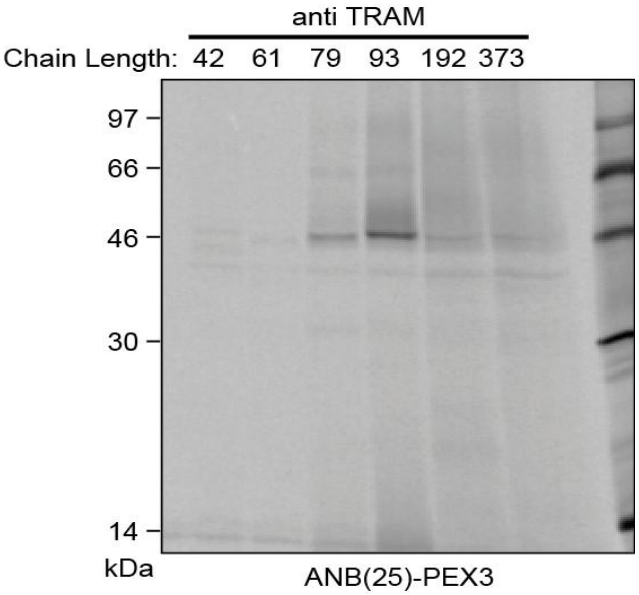


Figure S3 (related to main Figure 3). Photocrosslinking of PEX3 to TRAM depends on nascent chain length

[^{35}S]Met-labelled integration intermediates containing ANB(25)-PEX3 nascent chains were prepared in parallel in wheat germ extract (supplemented with canine ER microsomal membranes and 40 nM canine SRP) with lengths of 42, 61, 79, 93, 192, and 373 (full-length) residues. After photolysis, photoadducts were immunoprecipitated with antibodies directed against TRAM and analyzed by SDS-PAGE and phosphorimaging.

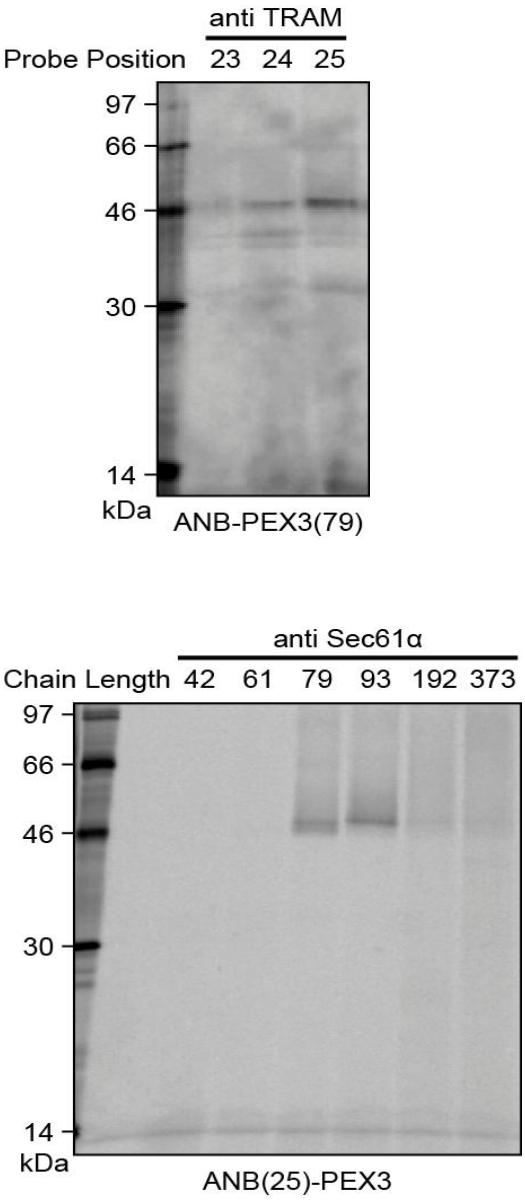


Figure S4 (related to main Figure 3). Uncropped phosphorimager scans of Figures 3D and 3E.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plasmids, mRNA, tRNA, SRP and Microsomes

All PEX3 constructs originated from the plasmid pcDNA3.1/PEX3mychis that encodes the human full length PEX3, as previously described (Kammerer et al., 1998). The introduction of a single amber stop codon at selected locations was done using the Quikchange protocol (Agilent Technologies). Bovine prolactin is encoded in the plasmid pSP64-BPL (Skach and Lingappa, 1993). For the membrane insertion of isolated PEX3-segments, HR1 (residues 14-36), HR2 (residues 108-130), or HR1-HR2 (residues 14-130) fragments were independently amplified and introduced into the modified *E. coli* leader peptidase sequence from the pGEM1 plasmid (Martinez-Gil et al., 2010) using the *Spe*I/*Kpn*I sites. The primary sequence of each construct was confirmed by DNA sequencing. mRNA was transcribed *in vitro* using SP6 RNA polymerase and PCR-generated DNA fragments of the desired length as before (McCormick et al., 2003). Reverse primers either contained an ochre stop codon to obtain full-length PEX3 translation products (e.g., for the budding assay) or lacked a stop codon for the generation of RNCs. Primer sequences are available from the authors on request. [^{14}C]Lys-tRNA^{amb}, εANB-[^{14}C]Lys-tRNA^{amb}, εNBD-[^{14}C]Lys-tRNA^{amb}, column-washed canine rough microsomes (CRM), and purified signal recognition particle (SRP) from dog pancreas in SRP buffer [50 mM triethanolamine (pH 7.5), 600 mM KOAc (pH 7.5), 6 mM $Mg(OAc)_2$, 1 mM DTT] were obtained from tRNA Probes, College Station, TX. SRP concentration was determined using ε_{280nm} = 1.0 x 10^6 $M^{-1}cm^{-1}$.

Cell-Free Translation in Rabbit Reticulocyte Lysate

In vitro translation of purified mRNA (typically 25 µl, 30°C, 40 min) was performed in the presence of rabbit reticulocyte lysate (Promega), [^{35}S]Met (0.4 µCi/µl), and, when

indicated, 4 equivalents (eq; Walter and Blobel, 1983) CRM. After translation, samples were either analyzed directly by SDS-PAGE and phosphorimaging (Bio-Rad FX), or membranes were collected by sedimentation (Beckman TLA100 rotor; 100,000 rpm; 5 min; 4°C) through a 0.5 M sucrose cushion in buffer A [30 mM HEPES (pH 7.5), 120 mM KOAc, 3.2 mM $Mg(OAc)_2$]. For proteolysis experiments, samples were treated with 4 mg/ml (Figure 2B) or 200 µg/ml (Figure 2E) proteinase K for 30 min on ice followed by the addition of 1 mM phenylmethylsulfonyl fluoride (PMSF). For carbonate extraction (Fujiki et al., 1982), membranes were incubated in carbonate buffer [0.1 M Na_2CO_3 (pH 11.5)] for 15 min on ice, centrifuged (Beckman TLA100 rotor; 100,000 rpm; 5 min; 4°C), washed, and resuspended in carbonate buffer. The supernatant and pellet fraction were neutralized with glacial acetic acid and further analyzed as above.

Photocrosslinking and Immunoprecipitation

In vitro translations (typically 50 µl, 26°C, 40 min) of truncated mRNAs were performed in wheat germ cell-free extract (tRNA Probes) in the presence of 40 nM canine SRP, 8 eq. CRM, [^{35}S]Met (1.0 µCi/µl), 0.6 pmol/µl [^{14}C]Lys-tRNA^{amb} / εANB-[^{14}C]Lys-tRNA^{amb} as indicated, and other components as described (Crowley et al., 1993). Samples were photolyzed on ice for 15 min using a 500W mercury arc lamp (McCormick et al., 2003). After photolysis, samples were collected by sedimentation (5 min for CRM or 60 min for free RNCs) through a 0.5 M sucrose cushion in buffer A as described above. Pellets were resuspended in 3% (w/v) SDS and 50 mM Tris-HCl (pH 7.5), then incubated at 55°C for 30 min. Samples were brought up to 500 µl with either buffer S [140 mM NaCl, 10 mM Tris-HCl (pH 7.5), and 2% (v/v) Triton X-100] for Sec61α-specific antibodies, or buffer T [150 mM NaCl, 1 mM EDTA, 50 mM Tri-HCl (pH 7.5), 1% (v/v) Triton X-100] for TRAM- or SRP54-specific antibodies.

Samples were precleared by rocking with protein A-Sepharose (Sigma-Aldrich; 40 µl; pre-equilibrated in buffer S or T) at 4°C for 1 h. After removal of the beads by centrifugation, the supernatants were incubated overnight at 4°C with affinity-purified rabbit antisera specific either for Sec61α or TRAM (McCormick et al., 2003), or for SRP54 (BD Biosciences). Protein A-Sepharose (40 µl, pre-equilibrated with buffer S or T) was then added and incubated for 4 h at 4°C. Sepharose beads were harvested by sedimentation and washed twice with 750 µl of buffer S or T, followed by a final washing in the same buffer without detergent. Samples were then analyzed by SDS-PAGE and phosphorimaging.

Fluorescence Spectroscopy

In vitro translations (500 µl total volume, 26°C, 40 min) of truncated mRNAs were performed in wheat germ cell-free extract in the presence of 0.6 pmol/µl εNBD-[^{14}C]Lys-tRNA^{amb} and other components as described (Crowley et al., 1993). To correct for the significant background signal due to light scattering from the ribosomes, equivalent blank translation reactions lacking NBD were prepared in parallel with [^{14}C]Lys-tRNA^{amb}. RNCs were purified by gel filtration at 4°C using a Sepharose CL-6B column (1.5 cm inner diameter x 20 cm) and buffer A as elution buffer. A slow flow rate was used during gel filtration to ensure the removal of noncovalently bound fluorophores. The absorbance at 260 nm of each 550 µl fraction was used to identify those fractions containing RNCs that elute in the void volume, and only the leading half of the void volume peak was pooled. After gel filtration, the absorbance at 260 nm of the two parallel samples (one with and one without NBD) was equalized before initiating spectral measurements. Steady-state fluorescence measurements were made with either an SLM-8100 or a Spex Fluorolog-3 spectrofluorometer at 4°C as described previously (Flanagan et al., 2003). Samples

(250 µl) were placed in 4 x 4 mm quartz microcells that were coated with phosphatidylcholine vesicles to minimize protein adsorption (Ye et al., 1991). The cuvette chamber was continuously flushed with N_2 to prevent condensation of water on the microcells. Emission intensity (λ_{ex} = 468 nm) was scanned at 1-nm intervals between 500 and 580 nm. Samples of purified RNCs with or without NBD in buffer A were titrated at 4°C by the sequential addition of known amounts of SRP in small volumes. After each addition, the emission intensities of the NBD and blank samples were measured after reaching equilibrium. After blank subtraction and dilution correction, the observed change in net NBD emission intensity (ΔF; λ_{ex} = 468 nm; λ_{em} = 528 nm, bandpass 4 nm) at each point in the titration was compared with the initial intensity (F_0) of the sample in the absence of SRP.

Budding Assay

Purified full-length PEX3 mRNA was translated in rabbit reticulocyte lysate in the presence ER microsomes as described above. The translation (60 min, 30°C) was stopped by addition of puromycin (2 mM final, 20 min, 4°C) and microsomes were collected by centrifugation through a 0.5 M sucrose cushion in buffer A as above. Membranes were incubated in 2.5 M urea in buffer A for 10 min at 4°C to remove peripherally bound PEX3 molecules. Membranes were collected by medium-speed (MS) centrifugation (20,000 g, 10 min, 4°C), washed once in urea buffer, and finally washed in buffer A. Such PEX3 containing donor membranes were resuspended in buffer A, and incubated with either rabbit reticulocyte lysate (the lysate was diluted to 60% of its original concentration in the budding reaction) or an equivalent amount of buffer A. Budding reactions also contained 2 mM puromycin and either an energy generating system (final concentrations: 16 mM phosphocreatine, 2 mM ATP, 2 mM GTP, 0.016 U/µl phosphocreatine kinase) or 1 U/µl apyrase. After incubation of the

budding reaction for 60 min at 30°C, donor membranes were removed by MS centrifugation, and the supernatant was analyzed by SDS-PAGE and phosphorimaging. In certain cases, the supernatant of a budding reaction was further subjected to high-speed (HS) centrifugation (Beckman TLA100 rotor; 55,000 rpm; 30 min; 4°C), and the pellet was resuspended in either carbonate buffer or 0.25 M sucrose in buffer A in the presence or absence of 1% (v/v) Triton X-100. After a second HS centrifugation, the protein content of the supernatant and pellet fractions was analyzed as above.

SUPPLEMENTAL REFERENCES

Crowley, K.S., Reinhart, G.D., and Johnson, A.E. (1993). The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. Cell *73*, 1101-1115.

Flanagan, J.J., Chen, J.C., Miao, Y., Shao, Y., Lin, J., Bock, P.E., and Johnson, A.E. (2003). Signal recognition particle binds to ribosome-bound signal sequences with fluorescence-detected subnanomolar affinity that does not diminish as the nascent chain lengthens. J. Biol. Chem. *278*, 18628-18637.

Fujiki, Y., Hubbard, A.L., Fowler, S., and Lazarow, P.B. (1982). Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. J. Cell Biol. *93*, 97-102.

Kammerer, S., Holzinger, A., Welsch, U., and Roscher, A.A. (1998). Cloning and characterization of the gene encoding the human peroxisomal assembly protein Pex3p. FEBS Lett. *429*, 53-60.

Martinez-Gil, L., Johnson, A.E., and Mingarro, I. (2010). Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. J. Virol. *84*, 5520-5527.

McCormick, P.J., Miao, Y., Shao, Y., Lin, J., and Johnson, A.E. (2003). Cotranslational protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. Mol. Cell *12*, 329-341.

Skach, W.R., and Lingappa, V.R. (1993). Amino-terminal assembly of human P-glycoprotein at the endoplasmic reticulum is directed by cooperative actions of two internal sequences. J. Biol. Chem. *268*, 23552-23561.

Walter, P., and Blobel, G. (1983). Preparation of microsomal membranes for cotranslational protein translocation. Methods Enzymol. *96*, 84-93.

Ye, J., Esmon, N.L., Esmon, C.T., and Johnson, A.E. (1991). The active site of thrombin is altered upon binding to thrombomodulin. Two distinct structural changes are detected by fluorescence, but only one correlates with protein C activation. J. Biol. Chem. *266*, 23016-23021.

RESULTS AND DISCUSSION

One of the main purposes of this thesis was to contribute to widen our current knowledge of insertion and assembly of membrane proteins that have intrigued our lab during the last 15 years. During these four years, the research associated to this work generated four published papers and one manuscript that is currently under revision, altogether constitute the four chapters of this thesis.

The results of this thesis were obtained using some of the main and principal techniques of molecular biology; clonation and site-directed mutagenesis, overexpression of chimeric proteins into heterologous systems, transcription/translation expressions *in vitro,* and photocrosslinking of truncated proteins into mammalian membranes.

All the techniques mentioned above lead us to perform experiments to achieve the main objective of this thesis; a better understanding of how TM helices inserts into the ER membrane and how they interact with each other to be stabilized and perform their diverse functions into the lipid environment.

The next pages are a summary of the results and a brief discussion of them, gathering full results and their correspondent discussion in four chapters.

R.1. Chapter One:

Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M.A., and Mingarro, I. *(2011). N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon.* ***Protein Science. 20, 179–186****.*

The objective of this paper, was to develop a new strategy to map membrane protein topology along protein biosynthesis, by determing whether glycosylation efficiency is affected by the position or the sequence of the glycosylation acceptor site. To this end, we started by generating truncated proteins that included *N*-linked glycosylation acceptor sites at different distances from the C-terminus. We utilized the well-characterized *Escherichia coli* inner membrane protein leader peptidase (Lep) harboring an Asn-Ser-Thr sequon, which is a well-known glycosylation motif. It has long been reported that the tripeptide

sequon Asn-Xaa-Thr is more efficiently glycosylated than Asn-Xaa-Ser (Gavel and Heijne, 1990); in fact, the occurrence rate of the former is about one third higher than that of the latter (39,161 and 30,579 sequons in our database, respectively), which is in agreement with a previous statistical survey (Ben-Dor et al., 2004).

We first determined that truncated Lep proteins that carried a C-terminal glycosylation tag were cotranslationally glycosylated. We found that translation of truncated Lep sequences with a glycosylation tag (**NST**MMS) in the presence of rough microsomal membranes (RM) was associated with an increase in the molecular mass, indicative of protein glycosylation. Furthermore, the increase in mass was corroborated by endoglycosidase H treatment (Endo H), a glycan-removing enzyme. Notably, when microsomal membranes were included post-translationally, after translation inhibition with cycloheximide, the C-terminal acceptor site was not glycosylated; suggesting that the truncated Lep protein is integrated into the membrane cotranslationally, via the ER translocon.

Next, we found that the glycosylation efficiency increased gradually with the distance between the acceptor site and the C-terminus. When the C-terminal glycosylation tag only included the three amino acid residues that formed the acceptor sequon (**NST**, 3 residues tag), the truncated polypeptide remained non-glycosylated. Extending the C-terminal tag to four residues (**NST**M) slightly increased glycosylation (~ 20%), and a C-terminal tag with five residues (**NST**MM) nearly doubled the glycosylation efficiency. Further extensions of the tag length rendered similar glycosylation levels (~50%). It should be noted that the methionine residue following the glycosylation sequon conferred optimal glycosylation efficiency when Thr was present at the hydroxyl (third) position (Mellquist et al., 1998). These results are consistent with an earlier study on truncated Lep, when the glycosylation acceptor site was placed six residues upstream of the stop codon (Nilsson and Heijne, 2000). Similarly, it has been shown that the recombinant mammalian concentrative Na^+-nucleoside cotransporter (rCNT1) could be glycosylated in *Xenopus oocytes* at an asparagine residue located six residues upstream of the C-terminal end (Hamilton et al., 2001).

We further compared these results with native glycoproteins, performing a statistical analysis using the sequences and their annotations from the UniProt database (http://www.uniprot.org, release 2010_09). After selecting nonredundant *N*-glycosylated proteins, the complete set of putative *N*-glycosylation sites was obtained by selecting only Asn-Xaa-Thr sequons. The final dataset contained 39,161 sequons of which 5,753 were experimentally validated. Native glycosylated sites located at the C-terminal regions were more prominent (13 occurrences) for sequons with the Asn amino acid located six residues upstream from the C-terminus. Nevertheless, the total number of glycosylation sites at this position relative to the total sequons (5,753 in native sequences) suggests that protein glycosylation near the C-terminus is a relatively rare event and explains the low glycosylation efficiency (~ 50%) in our experiments. As a consequence of the last observation, the tag with six amino acid residues (**NST**MMS) was selected for further experiments. Statistical studies have also shown that the glycosylation sequon occurs at the C-terminal end of well-defined glycoproteins at a lower frequency than that expected based on random chance (Ben-Dor et al., 2004; Gavel and Heijne, 1990) and that when *N*-glycosylation sites are contained within more than one extracytosolic loop, only the first loop is modified (Landolt-Marticorena and Reithmeier, 1994). Furthermore, those studies found that the glycosylation efficiency for Asn-Xaa-Thr sequons dropped when located close to the C-terminal end of the protein (Ben-Dor et al., 2004). Our results pointed out that this effect was emphasized at the very end of the protein. In fact, we found that at least six residues long C-terminal glycosylation tags were needed to achieve significant glycosylation; this validated their utility as a molecular tool in membrane protein topological studies.

r.1.1. Influence of the preceding amino acid residue on the glycosylation efficiency

We translated 20 variants of C-terminal-tagged truncated Lep proteins to examine systematically the preference in amino acid residues preceding the acceptor Asn for glycosylation efficiency. The probability of each amino acid residue preceding a verified glycosylation site has been calculated for the Asn-Xaa-Thr sequons in our nonredundant dataset. All 20 amino acids can be found preceding the Asn residue of the sequons, although significant differences between their probabilities occur in the experimentally

validated glycosylation sites. Methione and tryptophan displayed the lower probabilities. One explanation for this observation might be that the bulky side chains of these residues may block accessibility to the OST active site or the lipid carrier donor; another explanation could be that it may induce an unfavorable local protein conformation. Previous studies revealed that glycosylation was strongly inhibited when Trp was placed at the central Xaa position (Kasturi et al., 1997; Shakin-Eshleman et al., 1996), and it was somewhat inhibited when Trp followed the glycosylation sequon (Mellquist et al., 1998).

Experimentally, we found that the glycosylation efficiency of the NST sequon was also significantly lowered when it was preceded by Met, Trp, or Arg residues, which correlates with the results of our statistical analysis, especially in the case of Met and Trp. As expected, when a Pro residue preceded the acceptor Asn, glycosylation was significantly inhibited. However, Pro had a stronger inhibitory effect when it was located either at the central Xaa position (Shakin-Eshleman et al., 1996) or following the glycosylation sequon (Mellquist et al., 1998). It is interesting to note that Pro was never found preceding an experimentally verified glycosylation site in our database when the glycosylated Asn residue in the NXT sequon is located at six residues from the C-terminal end. However, this inhibitory effect was not observed when the Pro residue was inserted just before the acceptor Asn in a full-length Lep construct. This suggested that the residue preceding the glycosylation sequons only impacted glycosylation efficiency when the acceptor Asn residue was close to the end of the polypeptide. Indeed, of the 42 sequons Asn-Xaa-Thr located within the last eight residues, only two were preceded a Pro residue.

Our results also pointed out some average effect caused by the presence of acidic residues immediately before the glycosylation site. It is interesting to note that it has been described a notable reduction in the probability of finding acidic residues preceding the glycosylation site (Petrescu et al., 2004). Even more, this is accompanied by an increase probability of finding acidic residues preceding unoccupied glycosylation sites (Ben-Dor et al., 2004). However, both Asp and Glu have been found as average preceding NXT acceptor sites. The apparent discrepancies between our and the previous studies could arise from the fact that the later surveys included the glycosylation sites with Ser in the third position (NXS/T), whilst our database focused in NXT glycosylation sites. The other amino

acid residues appeared to have only minor effects on glycosylation efficiency; however, the Gly residue consistently induced higher glycosylation levels, and again, an increased probability of finding Gly preceding glycosylation sites was observed in our statistical analysis. We assumed that the flexibility that Gly confers on the conformation of the polypeptide chain might provide an advantage for OST catalysis. In fact, the structural conformation of the local region around the glycosylation sequon also influenced its accessibility and, consequently, its site occupancy. This supports the hypothesis that unfolding or flexibility is required for protein domains to be efficiently glycosylated. It should be also noted that there is a marked preference for hydrophobic amino acids immediately preceding the glycosylation site in our experimental data, which nicely correlates with previous (Petrescu et al., 2004) and the present statistical analysis of glycan-protein linkage, especially in the case of Leu that has been also found prevalent in glycosylated sequons.

Interestingly, statistical analyses of active bacterial *N*-glycosylation site consensus sequences showed that Asn, Phe, Ser, and Leu residues frequently precede the acceptor Asn (Kowarik et al., 2006). In the results presented, we found that these same residues and Gly were the best suited for glycosylation in a eukaryote OST. Based on these results, we propose that bacterial and eukaryotic systems might require similar sequences flanking the acceptor sites to adopt an optimal conformation upon the binding of OST.

Our findings about the efficiency of the *N*-glycosylation by the position of the motif have been refered in several recient papers papers (Gurba et al., 2012; Igura and Kohda, 2011; Nagaoka et al., 2012; Thaysen-Andersen and Packer, 2012). Particularly, Gurba et al, found that one sequon with Arg at the -1 position are considerably less glycosylated than sequons with Gly, as our study demonstrated.

r.1.2. Using C-terminal tags as a topological reporter for membrane protein topology studies

The final objective and main result of this chapter, was the validation of our experimental approach (introduction of *N*-glycosylation site by C-terminal tags) as a tool for topological studies of membrane proteins and their biogenesis intermediates. To prove our approach,

we fused the N-terminus of bacteriorhodopsin (bR) (from Trp10 to Val101) at the C-terminus of the engineered Lep sequence. We chose bR because it is a membrane protein with a well-defined topology, in which the N-terminus faces the extracellular side similarly to our chimeric constructs. *In vitro* transcription/translation of protein truncates using a glycosylable C-terminal tag after bR helix A (the first TM segment) rendered singly-glycosylated forms, indicating the insertion of bR helix A. Truncated polypeptides, which include the first two TM helices of bR, were efficiently doubly-glycosylated, demonstrating translocation of the glycosylation site included as a C-terminal tag, and validating our experimental approach.

Taken together, the results included in this first paper provided a rapid and efficient method for the determination of membrane protein topology that has been already adopted by other authors.

R.2. Chapter Two:

Martinez-Gil, L., **Bano-Polo, M.,** Redondo, N., Sánchez-Martínez, S., Nieva, J.L., Carrasco, L., and Mingarro, I. (2011*). Membrane Integration of Poliovirus 2B Viroporin.* ***Journal of Virology 85, 11315–11324****.*

Bano-Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J.L., Elofsson, A., and Mingarro, I. (2013). *Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion.* ***Journal of Molecular Biology 425, 830–840.***

Once we developed a tool that allow us to perform topological studies of membrane proteins, in this second chapter, that includes two articles, we have studied the topology of the Poliovirus 2B protein and unraveled the interactions that maintain its membrane embedded structure.

r.2.1. Viroporin 2B is an integral membrane protein.

Computer-assisted membrane protein topology prediction analysis is a useful starting point for experimental studies of membrane proteins. We parsed viroporin 2B amino acid sequence to test the performance of 10 popular prediction methods and found large discrepancies between their predictions. Two of the algorithms failed to predict 2B as an integral membrane protein, while other two of them assigned two TM segments for the protein. It should be noted that the reliability of a topology prediction can be estimated by the number of prediction methods that agree. Since six of the algorithms predicted 2B as a membrane protein with only one TM segment, these results clearly highlight that the presence of putative helical hairpin structures, which was not detected even by the methods predicting reentrant loops (MEMSAT-SVM and SPOCTOPUS), may be missed by current predictive methods, as previously suggested for a different TM helical hairpin (Nagy and Turner, 2007).

First, we determined the fused association of Poliovirus 2B protein to the membrane, by testing the resistance of the chimera 2B/P2 to alkaline extraction, observing an strong association to the membrane. Since this treatment disrupts microsomal membranes and releases any soluble luminal protein, this result indicates that the fusion protein is not translocated to the lumen of the microsomes. Urea treatments solubilized our fusion protein, indicating that secondary and tertiary structures in 2B play an important role in 2B insertion. The latter results contrast with previous work that showed that coxsackievirus 2B/enhanced green fluorescent protein fusions were resistant to urea extraction (de Jong, 2002). This discrepancy could be derived either from the differences found in the amino acid sequence of both 2B proteins or from the use of different fusion proteins in both cases. In fact, a significant influence of the P2 domain can be observed in our Triton X-114 partition experiments. Fusions containing the full P2 domain significantly partitioned into the aqueous phase, whereas the fusion of the only 50 N-terminal residues from this domain promoted the partitioning of this shorter chimera into the organic phase. These results were further corroborated by partition experiments using full-length 2B protein.

r.2.2. Insertion of the viroporin 2B hydrophobic regions into biological membranes

By challenging the hydrophobic regions of 2B in a model protein (Lep), we demonstrated that these regions do not span biological membranes when expressed separately, as expected according to their predicted apparent free energy of insertion. It should be mentioned that, in these Lep-derived constructs, the HR1 segment is forced to insert into the membrane with an N_{lum}/C_{cyt} topology, and this topological effect can prevent the proper TM disposition of this region. In fact, using a Lep-derived variant (Lep´), we demonstrated the TM disposition of HR1 when expressed with an N_{cyt}/C_{lum} orientation.

It has been shown previously that in some cases, a neighboring TM helix can promote membrane insertion of a marginally hydrophobic TM region (Enquist et al., 2009; Hedin et al., 2010; Ojemalm et al., 2012; Tamborero et al., 2011) and that there is a correlation between the polarity of a TM helix and its interaction area with the rest of the protein. Overall, the glycosylation pattern found in this system when both regions were expressed in-block (including their native turn region) suggests that, in the context of Lep-derived model proteins, the two hydrophobic regions of the viroporin 2B can insert into the membrane when expressed as an α-helical hairpin. These results further emphasize the hairpin orient both the N-/C-termini towards the cytosolic side.

r.2.3. Full-length Viroporin 2B integrates into the ER membrane through the translocon with an N-terminal/C-terminal cytoplasmic orientation.

The microsomal *in vitro* system closely mimics the conditions of *in vivo* membrane protein assembly into the ER membrane, however the presence of fused domains can influence its membrane insertion capacity. Hence, we next sought to investigate whether HR1 also could direct the integration into the ER membrane of the native 2B sequence (i.e., in the absence of nonviral fused domains) through the translocon. Because *N*-glycosylation acceptor sites are absent from the viroporin 2B sequence, several modifications were prepared to determine the TM disposition of different 2B-derived proteins. We used the C-terminal tags (**NST**MMM) previously validated (Chapter 1) to map the topology of truncated versions of viroporin 2B.

We engineered a glycosylation site at the N-terminal hydrophilic region of the 2B sequence, and translate short truncated molecules that only included HR1. The lack of glycosylation at this N-terminal engineered acceptor site together with the efficient glycosylation observed only when using the C-terminal acceptor site, as proven after EndoH treatment, strongly indicates that HR1 in the viroporin context is acting as a noncleavable signal sequence and is properly recognized by the translocon machinery that inserted it into the membrane with its N-terminus facing the cytoplasm (N_{cyt}/C_{lum} topology).

Further addition of HR2 residues to this construct resulted in the cytoplasmic reorientation of the C-terminus. The glycosylation data obtained in the context of the parental 2B sequence using engineered and truncated proteins provided compelling evidence that HR1-HR2 also may integrate cotranslationally into the membrane in the absence of fused domains.

Furthermore, glycosylation mapping analysis, both in an *in vitro* microsomal system and in cultured cells, suggests that both the N and C termini of viroporin 2B protein reside on the cytosolic side of biological membrane.

Because of the low hydrophobicity and the relatively poor insertion propensity found in the Lep system for HR2, it is predicted that the 2B C-terminal region will be translocated into the ER lumen. However, 2B hydrophobic regions also are responsible for the membrane anchoring of the virus 2BC polyprotein precursor, which performs its enzymatic and RNA binding activities in the cytosolic compartment. Thus, we speculated that some type of helix-helix interaction stabilizes the insertion of HR2 to keep an $N_{cyt}t/C_{cyt}$ 2B topology.

In our attempts to unravel the disposition of 2B in biological membranes, we designed a series of experiments where HR1 and various lengths of downstream sequence were translated as truncated proteins with a glycosylable C-terminal tag. A mutant polypeptide truncated at the end of the hydrophilic loop between HR1 and HR2 (60-mer) was highly glycosylated indicating that, the C-terminal glycosylation tag has been translocated into the lumen of the ER, and thus HR1 is integrated into the ER membrane in a N_{cyt}/C_{lum} orientation in this construct, as was observed when it was expressed in the Lep´

fusion. Extension of the 2B sequence to include roughly half of HR2 has a significant effect on this pattern (72-mer), suggesting some tendency of these truncated molecules to insert the C-terminus into the membrane. Moreover, extending the 2B sequence four residues (roughly one helical turn; 76-mer) substantially diminished glycosylation. The glycosylation level for the truncated protein shown it cannot be explained by an increased hydrophobicity of the added amino acids, since the residues ^{73}-CDAS-^{76} are notably polar. Interestingly, the addition of four leucine residues (ΔG^{pred} = 2.2 kcal/mol) instead of the ^{73}-CDAS-^{76} sequence [76-mer(L4)] strongly precludes glycosylation (less than 3%), demonstrating the clear hydrophobic effect of the leucine residues in this construct promoting insertion. This was further corroborated by extending the protein to include up to eight leucine residues [80-mer(L8)]. Finally, both the mutant truncated at the end of HR2 (80-mer) and the 2B full-length construct (2B/CtGlyc) had little effect on this pattern (~17 and ~10% glycosylation, respectively), indicating that the C-terminus of the great majority of these tagged proteins is cytosolic, and thus the HR2 sequence included in these constructs is integrated into the membrane in an N_{lum}/C_{cyt} orientation.

Next, we tested the topology of viroporin 2B in mammalian cells. To this end, 2B containing properly designed N-glycosylation sites at different positions were expressed in cultured cells, and their plasma membrane permeabilizing capacities assessed. No glycosylation of 2B was observed when this protein bears the *N*-glycosylation site at the amino, in the turn or the carboxy terminus of this viroporin. In addition, these two viroporin 2B variants retain their capacity to permeabilize cells to hygromycin B, suggesting that they are located at the membrane and exhibit the ability to alter membrane permeability. It should be noted that in the case of the *N*-glycosylation site located at the turn, although the location is luminal the absence of glycosylation is due to its proximity to the TM domains as previously reported (Nilsson and Heijne, 1993a; Orzáez et al., 2004). In conclusion, the *in vitro* and *in vivo* assays consistently indicate that both the N-and C-termini of viroporin 2B are facing the cytoplasm.

r.2.4. Helical hairpin insertion into the ER membrane can be driven by electrostatic interactions

A phylogenetic analysis of picornavirus 2B proteins has highlighted the presence of cationic residues in the first hydrophobic α-helix, and in the case of several genus the existence of an aspartic acid residue within the second hydrophobic segment (de Jong et al., 2008). In our attempts to identify possible helix-helix interactions that stabilize hairpin formation in biological membranes, we focused first on the aspartic residue mentioned above since this appeared to be a likely candidate to promote electrostatic interactions between the closely separated TM helices.

Once, demonstrated viroporin 2B topology using C-terminal tags, we extended this approach to establish the intramolecular requirements that maintain and promote α-helical hairpin insertion. We added a C-terminal glycosylation tag, and synthesized the first 76 residues of the PV 2B (76-mer), which contained the native aspartic residue (Asp74) to study the role of this polar residue.

To prove the role of Asp74 within these truncated protein forms, we replaced this residue by one or two lysine residues. The glycosylation level for these mutants increased significantly compared to the wild-type truncated sequence, suggesting that the presence of the lysine residues complicates membrane insertion of the second hydrophobic region. A similar effect was found when a second aspartic acid residue was included. The most likely explanation for the disruptive effect of this second aspartic acid residue (position 76) is that it lies on the side of the helix that faces the lipid (2 positions away from Asp74 or ~200º), imposing an unsurpassable penalty for membrane insertion of the second helix. Interestingly, replacement of Asp74 by glutamic residue rendered a glycosylation level similar to the native aspartic residue, indicating that a negative residue is necessary to maintain the stability of the α-helical hairpin.

r.2.5. Position-specific effects on helical hairpin formation by Lys–Asp pairs

We next focused on the positively charged residues present in the first hydrophobic region of the PV 2B sequence (residues 35-55). The protein sequence in this region contains three candidate lysine residues Lys39, Lys42 and Lys46. To find out which of these residues

would be involved in electrostatic interactions with the previously mentioned Asp74, we generated possible structural models using the Rosetta membrane protocol (Barth et al., 2009). These models clearly predicted a preference for Lys46 and Lys42 to interact with Asp74. Among the five largest structural clusters obtained, as compared to Lys39, interactions of Lys42 and Lys46 with Asp74 are 53 and 75 times more frequent respectively.

To prove experimentally the computational predictions, a series of experiments were designed, in which we separately replaced each lysine residues by an aspartic acid residue. The replacement of Lys42 by a negatively charged aspartic acid residue (K42D mutant) in truncated polypeptides (76 residues long, 76-mers) showed that Lys42 is involved in some electrostatic interactions between the two adjacent helices. In fact, these interactions can be partially restored as reported with the double mutant K42D/D74K. Thus, hairpin capacity for inserting into the ER membrane was almost unaffected by swapping charges between helices, which underscores the involvement of these residues in inter-helical electrostatic interactions. The same set of constructs was used to test Lys46 involvement in these electrostatic interactions, observing a similar effect when this lysine is modified.

To confirm that these interactions are relevant in the context of the 2B full-length protein, both Lys42 and Lys46 were separately mutated to aspartic acid and their membrane disposition experimentally determined. Both K42D and K46D mutants consistently showed a significantly increased glycosylation level, suggesting that the more efficient hairpin integration observed in the native 2B protein might be promoted by electrostatic interactions between Lys-Asp pairs placed in the adjacent TM segments.

Interestingly, it has been previously demonstrated that extramembranous positively charged residues placed near the cytosolic end of TM segments in membrane proteins promote membrane insertion of precedent hydrophobic helices (Lerch-Bader et al., 2008; Ojemalm et al., 2012). Inspection of PV 2B sequence reveals the presence of three positively charged flanking residues adjacent to the helical hairpin C-terminal end, which could then contribute to the higher integration efficiency for the full-length protein compared to truncated molecules where these flanking positively charged residues are absent.

r.2.6. Influence of turn residues on hairpin stabilization

Next we focused on the turn between the two TM segments as an additional source of stability of the helical hairpin motif. To investigate the role of the residues interconnecting both 2B TM segments we performed an alanine-scanning mutagenesis approach with the full-length Ct-tagged 2B protein. Alanine was used because it carries a non-bulky and chemically inert side chain, and more importantly, because it has a very low turn-promoting effect (Monné et al., 1999). The first two residues in the turn (Arg56 and Asn57) showed to be the more sensitive to alanine replacement. Interestingly, asparagine and arginine residues together with prolines have been found to behave as the strongest turn-promoting residues when placed in the middle of a 40-residue polyleucine segment (Monné et al., 1999). In addition, multiple sequence analysis of picornavirus 2B protein sequences highlighted Arg56 and Asn57 as fully conserved residues (Nieva et al., 2003), suggesting a relevant role in protein function. We also found that the contribution of turn residues can be combined with the interaction between charged residues appropriately located to stabilize the helical hairpin conformation by inter-helical hydrogen bonding, since the higher levels of truncated proteins glycosylation have been obtained when perturbed both turn and charged residues at the same time. Similarly, structural perturbations introduced by extracellular loop mutations in a cystic fibrosis TM conductance regulator (CFTR) appeared to affect hairpin conformation to a lesser extent in the context of more stable TM helix-helix interactions (Wehbi et al., 2007), highlighting the interplay between helix-helix interactions and turn contribution in modulating hairpin formation.

Furthermore, we collected all turns in helical hairpins from predicted integral membrane proteins of viral origin. A total of 3,654 such turns were found. Turn length and amino acid frequencies were calculated for these. We notice that 5-residue turn helical hairpins, like the one observed in PV 2B sequence, are among the most frequent in viral membrane, only surpassed by turns of length two.

Finally, we test the combined effect on the helical hairpin integration between the charge pairs found in the hydrophobic regions and the turn residues drives this process also in full-length proteins. To this end we designed double mutations including the more sensitive residue Asn57 (N57A), which appears to play an important role in hairpin

formation, and the lysines and aspartic acid residues involved in the electrostatic interactions found above (see Figure 23). The results obtained demonstrated an additive effect of the different protein domains in terms of helical hairpin stabilization.

r.2.7. α-helical hairpins: structural and biological relevance

Our observation that Lys-Asp pairs together with turn-promoting residues stabilize a TM helical hairpin conformation supports the folding of this domain at the initial stages of membrane protein biogenesis, i.e. within the translocon. In this regard, two TM segments have been found to accumulate within or adjacent to the translocon before final integration into the lipid bilayer (Pitonzo et al., 2009b; Saurí et al., 2005; 2007). Furthermore, molecular dynamics simulations of the translocon in a fully solvated lipid bilayer showed that two helices can coexist within the translocon (J. Gumbart, personal communication).

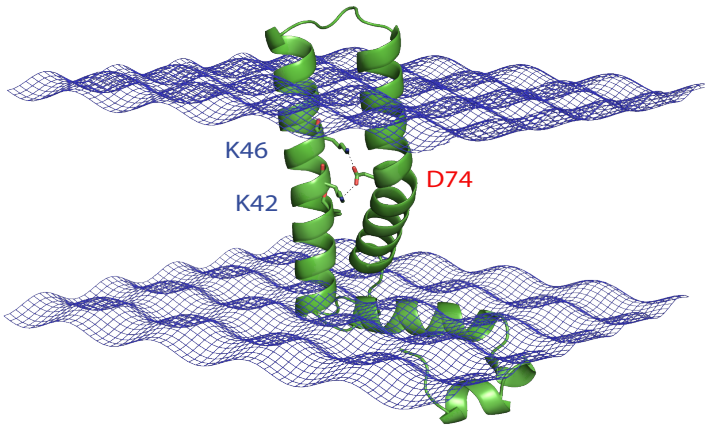


Figure 23. Model of full-length PV 2B in the context of a membrane.

An interesting possibility arising from these observations is that the translocon may facilitate membrane protein integration by allowing efficient polar interactions between consecutive TM segments harbouring charged residues, and thereby stabilizing them in the nascent polypeptide. In this particular case, the lysine residues located in the first TM segment may participate in keeping this domain in the translocon until the second TM segment reaches this location. If this were the case, helical hairpin formation inside the translocon might facilitate partitioning into the lipid bilayer by shielding the charged amino

acids that would otherwise constrain membrane integration efficiency. Although further studies are needed to unravel the details of this hypothesis, our results suggest that very specific helix-helix interactions can be formed within the context of the translocon and that such interactions together with turn constraints can have a dramatic effect into the poorly understood mechanism of multi-spanning membrane protein folding.

2.7.1. Biogenesis of viroporin 2B

Picornavirus 2B, a non-structural protein required for effective viral replication, has been implicated in cell membrane permeabilization during the late phases of infection. Three non-structural viral proteins, 2B, 2BC and 3A, have the capacity to alter cell membranes when individually expressed in cells (Aldabe et al., 1996; Barco and Carrasco, 1995; Doedens and Kirkegaard, 1995; Madan et al., 2010). The three genes encoding these proteins are clustered in the picornavirus genome, which is expressed in the form of a long polyprotein precursor that is proteolytically cleaved by virus-encoded proteases (Figure 24). The 2BC precursor is the most active protein in terms of membrane alteration. 2BC shows certain activities that are not observed when 2B and 2C are co-expressed (Aldabe et al., 1996). Nevertheless, several lines of evidence, including mutagenesis studies, suggest the capacity to enhance membrane permeability resides in the 2B moiety. Indeed, the synthesis of 2B alone disrupts membrane permeability in mammalian cells, although less efficiently than 2BC (Aldabe et al., 1996; Doedens and Kirkegaard, 1995). Perhaps the 2C protein serves to efficiently transport 2B to the plasma membrane. All three proteins, 2B, 2BC and 3A, enhance membrane permeability in prokaryotic cells (Lama and Carrasco, 1992). However, in mammalian cells it has not yet been possible to associate 3A with this effect (Aldabe et al., 1996). Perhaps, when individually expressed, this protein is retained in an intracellular compartment (Suhy et al., 2000).

Finally, 2B, 2BC and, in an even more active manner, 3A are all able to impair glycoprotein trafficking through the vesicle system (Aldabe et al., 1996; Doedens and Kirkegaard, 1995; Sandoval and Carrasco, 1997). A component of this system located at the ER may be the target for these proteins such that the Golgi apparatus becomes disorganized. The 2B product has been described to localize at the Golgi complex when expressed alone (de Jong, 2002).

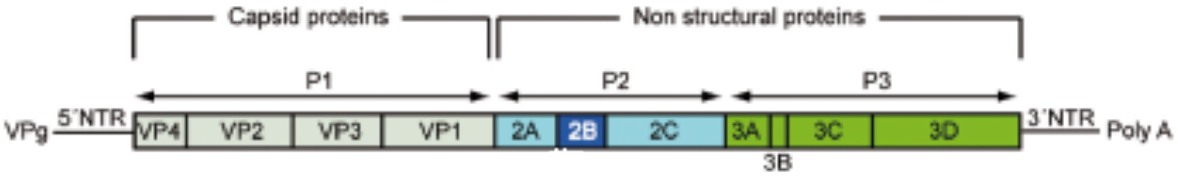


Figure 24. Non-structural Poliovirus 2B (A) Direct translation of poliovirus positive RNA gives rise to a polyprotein precursor, which is subsequently processed through proteolysis to produce three protein intermediates. 2B (dark blue) is one of the three non-structural regulatory products that originate from further processing of the second intermediate (top).

In the context of the PV P2 polyprotein, the membrane topology found in the present work leaves the protease cleavage sites of PV P2 facing the cytosol. In addition, this topology agrees with previous data obtained with different fusion proteins (de Jong, 2002). Furthermore, the localization of the C terminus at the cytosolic side of the membrane is consistent with the need for the proteolytic liberation of the 2B protein from the precursor 2BC polyprotein by a cytosolic viral protease cleavage (Foeger et al., 2002). Our findings further suggest a physiological role for translocon-mediated 2B integration into the ER membrane. Our combined analysis predicts a marginal propensity for 2B polypeptide to insert into membranes. On the other hand, upon viral entry, initially synthesized 2B or 2BC proteins will remain diluted in the cytosol of the infected cell. Marginal hydrophobicity together with low concentration are predicted to reduce the probability of the spontaneous insertion of 2B and 2BC into their primary target organelle: the Golgi complex (de Jong, 2002; Sandoval and Carrasco, 1997). Thus, we speculate that cotranslational insertion into the ER membrane is a particularly relevant phenomenon at the initial stages of the virus infectious cycle, during which both the viral mRNA levels and the concentration of the translated viral proteins are predictably low. Under those conditions, the cellular protein biosynthesis and vesicular transport machineries remain functional, and 2B and its 2BC precursor are likely synthesized as additional cell membrane proteins. From the ER, these proteins may reach the Golgi compartment to full regulatory and/or signaling functions that result in the disruption of Ca^{2+} homeostasis (Aldabe et al., 1997; de Jong et al., 2008) and vesicular transport inhibition (Barco and Carrasco, 1995; de Jong et al., 2008). At later stages, the canonical ER-Golgi protein-trafficking pathway no longer is functional and/or required for viral replication. The massive proliferation of cell endomembranes (viroplasm)

and the high levels of viral protein synthesis result in higher effective concentrations of these components inside the infected cell system. Under those conditions, it is predicted that the spontaneous insertion into the membrane of a significant amount of synthesized 2B and 2BC will ensue. In summary, viroporin 2B may use common structural arrangements to integrate into the ER membrane through the translocon, at least during the initial stages of the viral replicative cycle. The development of *in vitro* assays designed to dissect the membrane integration process will lead to a better understanding of the membrane permeabilization mechanisms that act during enterovirus infection.

2.7.2. Folding of viroporin 2B

Helical hairpins appear to be extremely common in multi-spanning integral membrane proteins, and is thus an essential structural motif to our understanding of MP folding and topology. In these structural motifs, the insertion of a polypeptide segment containing polar amino acids can be facilitated by the insertion of a closely spaced more hydrophobic region (Engelman and Steitz, 1981). Then, the membrane-buried polar groups are saturated with internal hydrogen bonds and salt bridges, and non-polar side chains are preferentially exposed on the bilayer-facing surfaces, balancing the physical constraints imposed by the non-polar core of the membrane bilayer and resulting in thermodynamic stability of the multi-spanning membrane protein.

In the hydrophobic environment of the membrane core, van der Waals helix-helix packing, hydrogen bonding and ionic interactions are the governing contributors to multi-spanning membrane protein assembly. These interactions can be modulated by the sequence context and the lipid bilayer properties (Cymer et al., 2012). Although hydrophobic interactions between TM segments are the more abundant, hydrogen bonding or salt-bridge formation between membrane-spanning charged residues are essential to drive membrane protein folding, while at the same time, reduces the unfavourable energetics of inserting charged residues into the membrane. Hence, interactions between polar residues in adjacent TM segments are proposed to favour membrane insertion.

R.3. Chapter Three

Bano-Polo, M., Baeza-Delgado, C., Orzáez, M., Marti-Renom, M.A., Abad, C., and Mingarro, I. (2012). *Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing.* ***PLoS ONE 7, e44263***

In the previous chapter, we found that membrane integration of a polypeptide segment containing polar amino acids can be facilitated by the insertion of a closely spaced more hydrophobic region, where polar residues present in both hydrophobic regions can interact to stabilize an α-helical hairpin. Interactions between polar residues in adjacent TM segments have been shown previously to favor membrane insertion (Buck et al., 2007; Meindl-Beinker et al., 2006; Zhang et al., 2007). The next objective, that constitute the third chapter of this thesis, has been to study the influence of polar residues in TM helix packing.

r.3.1. Presence of Ionizable amino acid residues in TM α-helices

Sequence analysis of high-resolution MP structures show that ionizable amino acid residues are present in TM helices, although at a low frequency (Baeza-Delgado et al., 2013). Insertion of these residues through the translocon has been proved to be feasible thanks to the overall hydrophobicity of the TM segment (Martínez-Gil et al., 2008) and depending on their position along the hydrophobic region (Hessa et al., 2007). In many cases, ionizable residues are involved in TM helix packing (Gratkowski et al., 2001; Hermansson et al., 2001; Zhou et al., 2001). Likely, hydrogen bonding (Gratkowski et al., 2001; Zhou et al., 2001) or salt-bridge (Chin and Heijne, 2000) formation with other membrane-spanning hydrophilic residues drives these interactions, while at the same time, reduces the unfavorable energetics of inserting polar or ionizable residues into the hydrophobic membrane core.

In this chapter, it has been analyzed the distribution of ionizable (Asp, Glu, Lys and Arg) amino acid residues in TM segments from high-resolution membrane protein structures, which have to energetically accommodate into the highly hydrophobic core of biological membranes by interacting favorably with its local environment. As expected,

ionizable residues (Asp, Glu, Lys, and Arg) are present at a low frequency level. All together, these residues constitute only 6.6 % of the residues within TM helices. Despite their lower presence, strongly polar residues are evolutionary conserved in TM proteins, which can be partially explained by their tendency to be buried in the protein interior and also in many cases due to their direct involvement in the function of the protein (Illergård et al., 2011).

Among the 792 TM helices included in our database, 366 helices (46.2%) contained at least one ionizable residue within the hydrophobic region (that is, the central 19 amino acid residues). Furthermore, 96 TM helices contained at least one acidic plus one basic residue in their sequence, and 20 of these helices present oppositely charged residues with the appropriate periodicity (*i*, *i*+4) to form intrahelical charge pairs. To gain more detailed insight into the structural role of these ionizable residues within the membrane core, we analyzed the environment of all these 20 helices. Approximately half of the ionizable residues (51%) found in these helices are buried in the protein interior, but the rest are partly exposed to the lipid face. Some of these lipid facing ionizable residues are located in pairs at the appropriate distance to form a salt-bridge.

r.3.2. Effects of ionizable residues on SDS-resistant TM helix packing

To investigate the effect of strongly polar residues in TM helix packing it has been used the GpA TM segment as a model (scaffold) segment. Polar mutations (T87D, T87K, I91D, and I91K) made on residues located at the helix-helix interface (Figure 25) abolished dimerization. Furthermore, it has been reported that T87S (which retains the side chain γ oxygen) permits dimer formation both in SDS micelles (Orzáez et al., 2005) and in *E. coli* membranes (Duong et al., 2007), whereas a bulkier hydroxylated side chain (T87Y) is strongly disruptive.

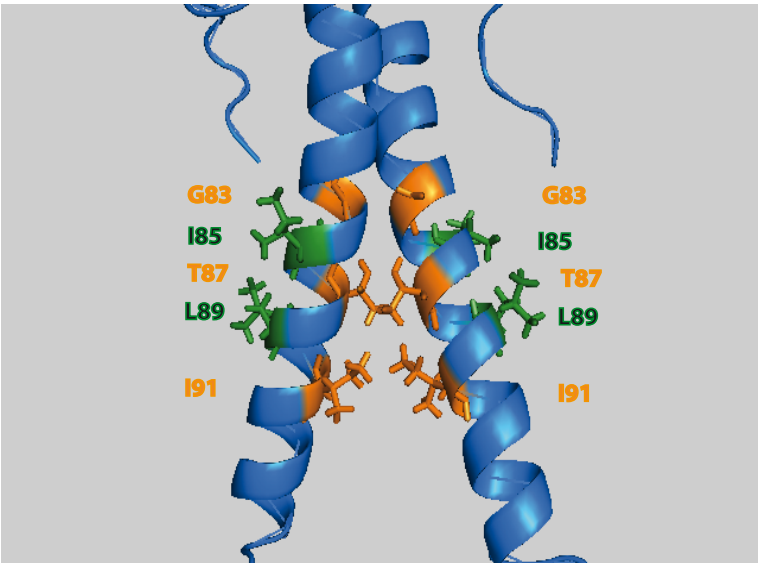


Figure 25. Structure of the glycophorin A TM homodimer. The side-chains of the five residues that have been studied are shown in orange for helix-helix interacting interfacial residues and in green for lipids facing residues, respectively.

However, point mutations corresponding to replacements of nonpolar residues located at the lipid-facing interface by ionizable residues gave rise to a more tolerated response. When Ile85 was substituted by ionizable side-chain residues, either negatively charged (I85D) or positively charged (I85K and I85R), the dimerization level was similar to native GpA TM sequence. It is commonly assumed that single ionizable residues should exist in their uncharged form within membrane-spanning helices (Herrmann et al., 2010). In fact, the pK_a values observed for Asp residues in hydrophobic helices were somewhat elevated (5-8.5) relative to those for Asp residues in solution (Caputo and London, 2003). Furthermore, the replacement of Leu89 by basic residues (L89K and L89R) had almost no effect, while its substitution by an acidic residue (L89D) abolished dimerization. The opposing consequences observed for Leu89 mutants can be explained taking into account the nature of the SDS-micelles used in these experimental conditions. These results suggest that L89D mutation alters the interaction of the protein with the negatively charged detergent micelle, possibly resulting in a structure that differs from a 'transmicellar' α-helix due to helix distortions and interaction with the polar micelle surface. This effect was not observed when the Asp residue was located in a more central position (I85D), where its carboxylate should be located away from the negatively charged sulfate groups of the SDS molecules. In this regard, the capacity of SDS to respond to such nuance of sequence in terms of SDS solvation of TM segments within protein-SDS detergent complexes has been

proved to be highly sequence (position) dependent (Tulumello and Deber, 2009). Nevertheless, the comparable electrophoretic migration observed for I85D and L89D suggests that the monomers associate with SDS quite similarly.

To identify the helix interface responsible of dimer formation in the Leu89 mutants, we designed double mutants that contained a non-polar highly disruptive mutation (G83L). Gly83 has been proved to be extremely sensitive, since all mutations tested disrupted the dimer completely (Lemmon et al., 1992). G83L mutant did not form any detectable dimer, and both double mutant proteins (G83L/L89K and G83L/L89D) containing this mutation did not dimerized, suggesting that the lysine residue introduced was not participating in the dimerization process, instead, the native dimerization motif is responsible of helix-helix interaction.

Given the 3.6-residue periodicity of an ideal α-helix, intrahelical charge pairs would be expected for (*i*, *i*+4) Lys-Asp pairs. To further assess if intrahelical charge pair formation can be tolerated in dimerizing TM sequences, we performed a double mutation combining two strongly polar dimerizing sequences (I85D/L89K), which only reduced dimerization by about 50% compared to the wild-type sequence. Similarly, I85K/L89K mutant retained the same level of dimerization, likely favored by a beneficial SDS solvation effect on the lysine residues (see above). On the contrary, when oppositely charged residues were located at the TM-interacting interface (T87D/I91K) dimerization was abrogated. Furthermore, when charge pairs include L89D mutation although facing the lipids, as for I85K/L89D, we found no evidence for dimer formation. These results suggest that charge pairs are tolerated only when located at the non-interacting interface, but solely at specific positions.

Recent mutational analysis of strongly self-interacting TM segments demonstrated that basic and acidic residues located at the helix-interacting interface participate in homotypic interactions (Herrmann et al., 2010). In this case, basic and acidic residues spaced (*i*, *i*+1) and (*i*, *i*+2) contribute to the interaction of model TM segments. To test this idea in the GpA sequence, we designed two mutants with appropriately spaced basic and

acidic residues (L89D/I91K and L90D/I91K), and no dimeric forms were observed in any of these proteins.

In light of our experiments in SDS micelles, we concluded that nonpolar to ionizable substitutions away from the dimer interface (lipid facing) in combination with N-terminal native GpA dimerization motif (including GxxxG sequence) does not perturb the dimerization process, while similar mutations positioned at the helix-interacting interface strongly compromise dimer formation.

r.3.3. Effects on insertion and packing into biological membranes

The next goal has been to test the molecular effect of the ionizable residues in biological membranes measuring changes in the insertion capacity of the GpA TM domain after introduction of ionizable residues at the more tolerant positions in terms of TM packing. The wild-type sequence of GpA TM segment efficiently inserts into the ER-derived microsomal membranes, while I85D mutation severely diminished membrane insertion capacity. On the contrary, L89K mutation allowed efficient insertion. The different effect observed for these two mutants can be explained by differences in amino acid side chain size and the position of the residue relative to the midpoint of the TM sequence. Hence, in the case of L89K, the longer side chain of this cationic amino acid and its proximity to the membrane interface compared to I85D may allow the hydrophilic moiety of the lysine residue to snorkel, that is, to approach its ε-amino group toward the interfacial and aqueous region, close to the negatively charged phospholipid head groups. Next, a construct with an Asp-Lys pair at the same positions (double mutant I85D/L89K) was glycosylated somewhat more efficiently than the I85D construct, supporting the idea that an intrahelical salt-bridge or hydrogen bond interactions between Lys and Asp side chains located on the same face of a TM helix can facilitate its insertion into biological membranes by reducing the free energy of membrane partitioning, as previously suggested in a similar system (Chin and Heijne, 2000).

Finally, the effect of ionizable residues in TM packing in bacterial membranes was assessed using the ToxCAT assay where TM-mediated dimerization of the chimera in the *E. coli* inner membrane results in transcriptional activation of a reporter gene encoding

chloramphenicol acetyltransferase (CAT), with the level of CAT protein expression indicating the strength/intensity of TM helix-helix interactions. Dimerization of wild-type and mutant sequences carrying ionizable residues was assessed along with a GpA point mutant (G83I) that disrupts homodimerization as negative control. The I85D mutant was found to dimerize in this system to about 35% of the level shown by wild-type GpA. Interestingly, L89D mutant, which precludes dimer formation in the presence of SDS micelles, appears to retain some dimerization capacity (21 ± 4%, normalized dimerization), which highlights the influence of the specific lipid environment during the assembly of TM segments (Martínez-Gil et al., 2011). Nevertheless, differences in TM segment length and flanking residues sequences may alter the dimerization process in the two systems, which are difficult to rationalize. Mutation of Leu89 to lysine (L89K) had a smaller effect on TM dimerization, and double mutant I85D/L89K still retained some dimerization capacity. In agreement with these data, recent molecular dynamics simulations suggested that a lysine residue outside the contact interface could exert a significant influence on TM helix association affinity of the bacteriophage M13 major coat protein because the extent of their burial in the membrane could be different in monomers and dimers (Zhang and Lazaridis, 2009). Together, our data indicate that the presence of ionizable residues does not preclude membrane insertion and allows dimer formation in bacterial cells.

R.4. Chapter Four

Mayerhofer, P.U., **Bano-Polo, M.,** Mingarro,I., Johnson, A.E. *Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co- translational Integration into and Budding Vesicle Exit from the ER Membrane* ***under revission***

In the fourth chapter, we studied the integration of the human peroxisomal membrane protein PEX3 into the ER membrane. Most of the results obtained were performed at the laboratory of the Professor Arthur E. Johnson, where I carried out a short stay of 15 weeks.

r.4.1. Ribosome-bound human PEX3 nascent chains are recognized and bound by mammalian SRP

As a prerequisite to understand the early stages of peroxisomal biogenesis, it is essential to ascertain how peroxins are inserted into the ER membrane. Two hydrophobic regions, HR1 and HR2 have been identified in human PEX3 (Kammerer et al., 1998). Since HR1 emerges first from the ribosomal exit tunnel during synthesis, HR1 interactions were examined by using environmentally sensitive probes. To investigate the targeting of PEX3, photoadducts were analyzed by immunoprecipitation using antibodies specific for SRP54, the signal sequence-binding component of SRP (Krieg et al., 1986). Since [^{35}S]Met-labelled 93 residues long nascent chains reacted covalently with SRP54, the photoreactive probe in HR1 was proved adjacent to SRP54. Thus, HR1 was recognized and bound to SRP as it emerged from the ribosome.

The HR1 interaction with SRP indicates that the nonpolar HR1 acts as a signal sequence. Then, we tested whether HR1 also function as a TM segment to anchor PEX3 in the membrane. PEX3 HR sequences were engineered into the *Escherichia coli* inner membrane protein leader peptidase (Lep) and the glycosylation pattern revealed that when analyzed in the Lep context HR1, but not HR2, was efficiently integrated into the ER membrane. Furthermore, carbonate extraction of PEX3 and a protein derivative lacking HR1 showed that HR1 is necessary and sufficient for stable insertion of PEX3 into the ER bilayer. The large C-terminal domain of ER-inserted PEX3 was therefore exposed to the cytosol, a topology previously described for peroxisomal-localized PEX3 protein (Kammerer et al., 1998). We conclude that the N-terminal hydrophobic HR1 of PEX3 acts as a non-cleavable signal-anchor TMS that is recognized by the SRP. These data provide the first direct evidence that the TMS of a peroxisomal (integral) membrane protein (PMP) can be recognized and bound by the SRP when it emerges form the ribosomal exit tunnel.

r.4.2. PEX3 is co-translationally integrated into the ER membrane at the translocon

To examine SRP dependence of PEX3 targeting to the translocon, PEX3-RNCs were translated in a wheat germ extract containing a low endogenous content of SRP that in this system ribosome nascent chain complexes (RNC) targeting to column-washed canine rough

microsomes (CRM) is dependent on added canine SRP (Krieg et al., 1986; Martínez-Gil et al., 2010; Tamborero et al., 2011). After photolysis and immunoprecipitation using antibodies specific for Sec61α, covalent photoadducts between Sec61α and PEX3 nascent chains were observed only in the presence of SRP. Thus, SRP is required to target nascent PEX3 to the translocon.

To further characterize HR1 interactions with translocon components, we used a high-resolution photocrosslinking approach. Photocrosslinking to translocon proteins was determined by immunoprecipitation with antibodies specific for Sec61α and TRAM. HR1 proximity to translocon proteins was then examined as a function of nascent chain length. Since RNCs 93-residues long with εANB-Lys at residue 25 photocrosslinked to both Sec61α and TRAM, RNCs with increasing nascent chain lengths were prepared in parallel, photolyzed, and analyzed by immunoprecipitation. When nascent chain length increased beyond 93 residues, HR1 was no longer adjacent to Sec61α. TRAM-containing photoadducts were observed with nascent chain lengths of 93 and 122, 148 to a lesser extent, and not at all for nascent chains 192 or more residues. Since HR1 was adjacent to TRAM, but not to Sec61α, at 122 residues, we concluded HR1 was retained close to TRAM longer than to Sec61α, consistent with earlier data showing a TM segment passing sequentially from Sec61α to TRAM during integration at the translocon (Do et al., 1996; Saurí et al., 2007). PEX3 therefore inserts co-translationally into the ER membrane via the standard SRP-dependent and translocon-mediated multistep pathway.

r.4.3. PEX3 exits the ER by ATP-dependent selective packaging into budding vesicles

Given the sub-stoichiometric number of SRPs relative to ribosomes (in yeast, there are 1-2 SRPs for every 100 ribosomes (Raue et al., 2007)), it is certainly possible that some PEX3 molecules may escape recognition by SRP and be inserted post-translationally into pre-existent peroxisomes (Matsuzaki and Fujiki, 2008) or the ER (Kim et al., 2006; Toro et al., 2009). In such cases, the initial co-translationally inserted PEX3 may then serve as docking factor for PEX19•cargoPMP complexes (Matsuzaki and Fujiki, 2008) in the ER and thereby concentrate PMPs or PMP sub-complexes in a spatially defined area of the ER (van der Zand et al., 2012).

Finally, we tested using a cell-free vesicle budding assay, recently established in yeast (Agrawal et al., 2011; Lam et al., 2011), if PMP-containing carrier vesicles are released from the ER in an ATP-dependent process. We found PEX3 packed into vesicles that bud from mammalian ER membranes when full-length PEX3 was translated *in vitro* in the presence of canine ER microsomes.

Our data therefore provide the first direct evidence that human PMPs are actively and selectively extracted from mammalian ER membranes in a cytosol-dependent and ATP-consuming vesicle budding reaction. Moreover, these data are consistent with small ER-derived vesicles playing a role in PEX3 trafficking to peroxisomes.

The combined data presented in this chapter establish that human PEX3 is targeted to mammalian ER membranes, integrated co-translationally at mammalian translocons, and then selectively packaged and extracted from the ER membrane, via an energy and cytosol-dependent budding reaction. By experimentally characterizing the entire pathway required for PEX3 passage through the ER membrane, from recruitment and entry to exit and discharge, the transient role of the ER in peroxisomal biogenesis is now defined from entry to exit.

CONCLUSSIONS

Chapter One:

Bano-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M.A., and Mingarro, I. *(2011). N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon.* ***Protein Science. 20, 179–186****.*

- Glycosylation efficiency depends on the distance between the acceptor Asn residue and the C-terminus of the polypeptide, increasing gradually with the distance between the acceptor site and the C-terminus.
- We found that at least six residues long C-terminal glycosylation tags were needed to achieve significant glycosylation
- The amino acid residue preceding the acceptor Asn affected glycosylation efficiency. Hence preceding Met, Trp, or Arg residues significantly lowered glycosylation efficiency.
- Statistically, all 20 amino acids can be found preceding the acceptor Asn residue of the sequons, although significant differences between their probabilities occur in the experimentally validated glycosylation sites.
- We validated the utility of C-terminal glycosylation tags in membrane protein topological studies, providing a rapid and efficient method for the determination of membrane protein topology.

Chapter Two:

Martinez-Gil, L., **Bano-Polo, M**., Redondo, N., Sánchez-Martínez, S., Nieva, J.L., Carrasco, L., and Mingarro, I. (2011). *Membrane Integration of Poliovirus 2B Viroporin.* ***Journal of Virology 85, 11315–11324.***

Bano-Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J.L., Elofsson, A., and Mingarro, I. *(2013). Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion.* ***Journal of Molecular Biology 425, 830–840.***

- The hydrophobic regions of 2B, do not integrate as isolated TM segments in the presence membrane of ER-derived membranes.
- The two hydrophobic regions of the viroporin 2B can insert into the membrane in the Lep system when expressed as an α-helical hairpin.
- Viroporin 2B is integrated cotranslationally through the ER translocon as a double-spanning integral membrane protein with an N-/C-terminal cytoplasmic orientation.
- HR1 and HR2 interact with each other to form a helix-turn-helix (helical hairpin) motif that traverses the lipid bilayer. An additional source of stability of this motif can be the turn residues located between the two helices.
- The same topology was adopted by functional viroporin 2B in cultured cells, where viroporin 2B membrane-permeabilizing capacities were maintained.
- Negatively charged residue (Asp74) turned to be critical for hairpin integration.
- Rosetta membrane protocol models clearly predicted a preference for Lys46 and Lys42 to interact with Asp74, which was further demonstrated experimentally.
- Native 2B protein folding might be promoted by electrostatic interactions between Lys–Asp pairs placed in adjacent TM segments.
- Our observations support helical hairpin formation at the initial stages of membrane protein biogenesis, that is, within the translocon.
- We propose that translocon may facilitate membrane protein integration by allowing efficient polar interactions between consecutive TM segments harboring charged residues and thereby stabilizing them in the nascent polypeptide. In this particular

case, the lysine residues located in the first TM segment may participate in keeping this domain in the translocon until the second TM segment reaches this location.

Chapter Three

Bano-Polo, M., Baeza-Delgado, C., Orzáez, M., Marti-Renom, M.A., Abad, C., and Mingarro, I. ***(2012a).*** *Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing.* ***PLoS ONE 7, e44263.***

- We analyzed the distribution of ionizable (Asp, Glu, Lys and Arg) amino acid residues in TM segments from high-resolution membrane protein structures. As expected, these residues are present at a low frequency level. All together, constitute only 6.6% of the residues within TM helices.
- In light of our experiments in SDS micelles, it can be concluded that nonpolar to ionizable substitutions away from the dimer interface (lipid facing) in combination with N-terminal native GpA dimerization motif (including GxxxG sequence) does not perturb the dimerization process, while similar mutations positioned at the helix-interacting interface strongly compromise dimer formation.
- An intrahelical salt-bridge or hydrogen bond interactions between Lys and Asp side chains located on the same face of a TM helix can facilitate its insertion into biological membranes by reducing the free energy of membrane partitioning.
- Our data indicate that the presence of ionizable residues does not preclude membrane insertion and allows dimer formation in bacterial cell membranes.

Chapter Four

Mayerhofer P.U., Bano-Polo et al**. "Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co-translational Integration into and Budding Vesicle Exit from the ER Membrane"* ***Under revision (2013)

- The first hydrophobic region of PEX3 (HR1) was recognized and bound by the SRP as it emerged from the ribosome, indicating that the nonpolar HR1 acts as a signal sequence.
- Isolated HR1, but not HR2, was efficiently integrated into the ER membrane. HR1 is necessary and sufficient for stable insertion of full-length PEX3 into the ER bilayer.
- Covalent photoadducts between Sec61α and PEX3 nascent chains were observed only in the presence of SRP. Thus, SRP is required to target nascent PEX3 to the translocon.
- HR1 was retained next to TRAM longer than to Sec61α, consistent with earlier data showing a TM segment passing sequentially from Sec61α to TRAM during integration at the translocon. PEX3 therefore inserts co-translationally into the ER membrane via the standard SRP-dependent and translocon-mediated multistep pathway.
- We found the first direct evidence that human PMPs are actively and selectively extracted from mammalian ER membranes in a cytosol-dependent and ATP-consuming vesicle budding reaction. This work defines the role of the ER in mammalian peroxisomal biogenesis from entry to exit.

-

RESUMEN

R.1. INTRODUCCIÓN

Todas las células, ya sean eubacterias, arqueas o eucariotas, están delimitadas por membranas, cuyo componente estructural básico es una fina capa de moléculas anfipáticas organizadas en dos monocapas lipídicas enfrentadas, en las que residen multitud de proteínas íntimamente asociadas a esta la bicapa de múltiples maneras. La membrana juega un papel fundamental para la supervivencia de las células, define sus límites estableciendo una barrera físico-química para moléculas polares a la vez que delimita los diferentes compartimentos celulares.

Actualmente se dispone de un elevado número de genomas secuenciados (4329, el 10 de mayo de 2013). El análisis del conjunto de las ORFs identificadas en distintos genomas nos indica que aproximadamente el 25% de dichas secuencias, independientemente del organismo al que pertenezcan, corresponden a proteínas de membrana (Krogh et al., 2001). En la actualidad, que está tan en boga el concepto de "ciencia translacional", cabe destacar que la industria farmacéutica se encuentra especialmente interesada en estas proteínas ya que se ha demostrado que numerosos receptores de membrana pueden ser empleados como potentes dianas terapéuticas, de hecho más del 50% de los fármacos actuales van dirigidos a este tipo de proteínas (Chin et al., 2002). A pesar de la importancia de las proteínas de membrana únicamente el 1.5% de las estructuras depositadas en el "Protein Data Bank" (PDB) (incluyendo las estructuras redundantes) corresponde a estructuras de polipéptidos pertenecientes a este grupo. Esto se debe, fundamentalmente, a las dificultades técnicas que presenta el trabajo con este tipo de proteínas, lo que convierte a las novedosas aproximaciones moleculares y a todos aquellos métodos computacionales que nos permitan obtener modelos estructurales de proteínas de membrana en grandes herramientas de trabajo, tanto por los propios resultados que ofrecen como por servir de punto de partida para el diseño de futuros experimentos.

Las proteínas de membrana, responsables de actividades celulares clave, se ensamblan mediante un proceso complejo que implica la síntesis de la proteína por

ribosomas asociados a membranas del retículo endoplásmico (ER) y unidos transitoriamente al complejo proteico del translocón. Este complejo multiproteíco proporciona un canal a través del cual se transfiere la proteína recién sintetizada a la bicapa lipídica. Después de completarse la síntesis, los ribosomas se separan del translocón y la proteína es integrada en la membrana donde finalmente adopta su estructura nativa. Independientemente de cómo llegue la proteína de membrana a la bicapa, una vez en ella tiene que residir en un estado de mínima energía. En principio, para estudiar cómo se alcanza este estado de mínima energía podría desplegarse al encontrarse en un entorno altamente hidrofóbico y plegarse en la bicapa. Sin embargo, esto es virtualmente imposible con proteínas de membrana, puesto que son insolubles en la bicapa en la forma desplegada debido al coste energético que supone la exposición del esqueleto polipeptídico (de naturaleza polar) al núcleo hidrocarbonado de la membrana. Además, las proteínas de membrana son insolubles en la fase acuosa tanto en su forma plegada como desplegada debido al carácter altamente hidrofóbico de la mayoría de los aminoácidos que las componen.

Básicamente han sido descritos dos motivos estructurales en las regiones transmembrana (TM): los barriles beta y los haces de hélices alfa. Estos tipos de estructuras secundarias permiten la inserción en la membrana de los esqueletos polipeptídicos al establecer puentes de hidrógeno intramoleculares y excluir el agua de solvatación del interior de las proteínas, reduciendo la polaridad intrínseca de los grupos CO y NH del enlace peptídico, cuyo momento dipolar permanente impediría su inserción en el núcleo hidrofóbico de la membrana. Las proteínas de membrana más comunes son aquellas basadas en motivos en alfa hélice, siendo el objeto de estudio en esta tesis. En lo que a las hélices alfa transmembranales se refiere, los residuos de aminoácidos que las forman se localizan en una distribución determinada dentro de la estructura de la bicapa dependiendo de sus características químico-físicas. Lo más frecuente es que en el núcleo hidrofóbico se encuentren aminoácidos hidrofóbicos (L, I, V, A, F, M) y que los residuos polares o cargados se ubiquen en localizaciones interfaciales o incluso extramembranosas, aunque existen ejemplos en los que fragmentos TM contienen aminoácidos cargados embebidos en el núcleo hidrofóbico de la membrana (Baeza-Delgado et al., 2013; Kauko et al., 2010).

Durante su biogénesis, tanto las proteínas integrales de membrana como aquellas solubles del espacio extracelular (proteínas de secreción) han de cruzar la membrana parcial o totalmente. Ambas emplean la misma maquinaria, el translocón, mencionada anteriormente, para su inserción y translocación. El translocón consiste básicamente en un canal que atraviesa la bicapa lipídica, el cual cumple dos funciones, por un lado permite que las proteínas solubles del lumen o de secreción atraviesen completamente la membrana del ER y por otro, que los fragmentos TM de las proteínas integrales de membrana se introduzcan lateralmente en la bicapa.

En este mecanismo, la translocación e inserción de la proteína en la membrana se produce en un proceso acoplado a la traducción. La síntesis de las proteínas comienza en el citoplasma y requiere el direccionamiento de los polipéptidos mediante la utilización de secuencias de señalización (SS), que consisten en uno o más residuos cargados seguidos de entre 12 y 20 aminoácidos hidrofóbicos. Cuando esta región emerge del ribosoma es reconocida por la partícula de reconocimiento de la secuencia señal (SRP) (Walter and Johnson, 1994), la cual es capaz de detener momentáneamente la traducción y dirigir el complejo (mRNA-ribosoma-polipéptido naciente) al translocón. En un primer paso el complejo es direccionado a la membrana del RE, proceso guiado por la afinidad entre la SRP y su receptor (SR), el cual se encuentra en las proximidades del translocón (Figura 26). A continuación y una vez anclado a la membrana a través del SR, el complejo es transferido al translocón (Halic and Beckmann, 2005).

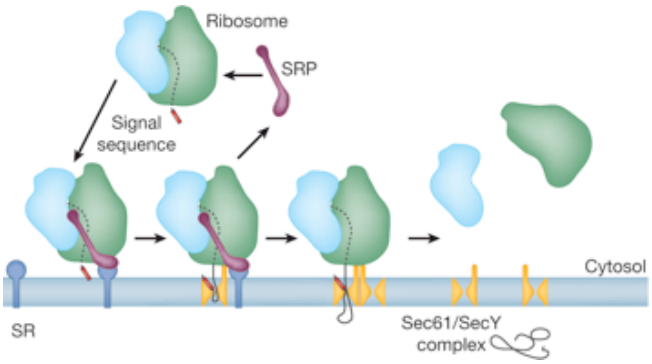


Figura 26. Inserción co-traduccional. La SRP, tras reconocer la secuencia señal, dirige el ribosoma con la cadena naciente y el mRNA al canal del translocón mediante la interacción con su receptor (SR). Es entonces, una vez se ha unido el ribosoma al translocón cuando la SRP se libera en un proceso dependiente de la hidrólisis de GTP y la traducción se reinicia. La cadena naciente es ahora capaz de pasar a través del translocón. Adaptado de (Rapoport, 2007)

A medida que las proteínas de membrana atraviesan el canal del translocón los segmentos TM son insertados en el interior de la bicapa lipídica a través de una apertura lateral del canal. Así, los segmentos suficientemente hidrofóbicos se integrarán en la membrana a través de esta apertura lateral una vez estén ubicados en la posición adecuada dentro del canal. El tamaño estimado de esta apertura parece indicar que los segmentos TM salen del canal de uno en uno (Heinrich and Rapoport, 2003) o como mucho en parejas (Saurí et al., 2005) aunque también existen evidencias experimentales de las inserción de varias hélices en bloque (Sadlish et al., 2005). En este sentido, el complejo del translocón permitiría interacciones entre diferentes dominios TM de una misma proteína antes de que el proceso de integración haya finalizado completamente, lo cual facilita la inserción en la bicapa de segmentos que, sin estas interacciones, no serían capaces de integrarse (White and Heijne, 2008).

La correcta topología de una proteína de membrana (determinación del número de segmentos TM y su orientación relativa en la membrana) es fundamental para poder llevar a cabo su función biológica, y en ausencia de información estructural de alta resolución, se convierte en una herramienta imprescindible para la comprensión de los sistemas de membrana. Generalmente la topología que adquiere una proteína es única, aunque se han identificado casos en los que una misma secuencia es capaz de insertarse en la membrana con dos topologías opuestas (Rapp et al., 2006; 2007). Existen diferentes factores que determinan la topología de una proteína, la presencia de secuencia de direccionamiento a membrana (SS), entre los que hay que destacar las cargas flanqueantes al los fragmentos TM, el plegamiento de dominios estructurales N-t y la hidrofobicidad del segmento TM.

R.2. OBJETIVOS

El objetivo general de esta tesis es la comprensión de la inserción y el ensamblaje en la membrana lipídica de α-hélices presentes en las proteínas de membrana. Durante el estudio intensivo de estas proteínas, en esta tesis se abordaron los siguientes objetivos concretos:

- Desarrollar una nueva estrategia molecular basada en la glicosilación co-traduccional para la determinación de la topología de proteínas de membrana.
- Analizar la topología en la membrana del retículo endoplasmático de la proteína viroporina 2B de *Poliovirus*. Describir las interacciones intramoleculares que estabilizan la estructura de la proteína en la membrana biológica.
- Estudiar el efecto de los residuos polares en la inserción y plegamiento de las α-hélices de las proteínas de membrana.
- Describir el papel del retículo endoplasmático en la biosíntesis de la peroxina PEX3 humana en el contexto de la biogénesis peroxisomal.

R.3. METODOLOGÍA

r.3.1. Predicción de fragmentos transmembrana

Históricamente se ha considerado que la predicción de dominios TM debería de ser un proceso relativamente sencillo, debido a las fuertes restricciones que impone la membrana sobre la composición de aminoácidos en los dominios insertados en la bicapa y sobre la estructura secundaria que éstos pueden adoptar. Actualmente, el incremento de estructuras de proteínas de membrana depositadas en el "Protein Data Bank" (PDB) muestra que existe una gran variabilidad en las regiones TM, lo que dificulta la identificación de los dominios TM.

Existen dos grandes grupos de métodos para la identificación de dominios TM. Las primeras aproximaciones, más simplistas, basadas en la identificación de regiones de entre 15 y 30 aminoácidos donde predominen los residuos hidrofóbicos. A medida que se han ido obteniendo más datos experimentales, las reglas que definen qué aminoácidos son permitidos en las membranas se han ido refinando (p ej. preferencia de los aminoácidos aromáticos por la zona de interfase de las bicapas o la mayor incidencia del efecto del *snorkel* los extremos de los dominios TM (Chamberlain et al., 2004)) y con ellas los métodos de predicción.

Los programas empleados en este trabajo fueron:

PHDhtm (Rost et al, 1996) http://www.predictprotein.org
SOSUI (Hirokawa et al, 1998) http://bp.nuap.nagoya-u.ac.jp/sosui/
Tmpred (Hofmann, 1993) http://www.ch.embnet/cgbin/TMPERD_from.htm
MEMSAT (Jones et al, 1994) http://bioinf.cs.ucl.ac.uk
MPEx (Jaysinghe S et al, 2006) http://blanco.biomol.uci.edu/mpex/
DAS (Cserzö et al, 1997) http://www.sbc.su.se/~miklos/DAS/
TMHMM (Krogh et al, 2001) http://www.cbs.dtu.dk/services/TMHHM
HMMTOP (Tusnady & Simon, 2001) http://www.enzim.hu/hmmtop/
TOPPRED(Claros&vonHeijne,1994)http://biowb.pasteur.fr/seqanal/tmp
ΔG prediction server (Hessa et al, 2005) http://www.cbr.su.se/DGpred/

Recientemente se ha implementado esté último algoritmo (*ΔG prediction server*) en el que dada una secuencia de aminoácidos se estima la diferencia de energía libre aparente (ΔG_{app}) para la inserción de dicha secuencia en la membrana a través del translocón, teniendo en cuenta la posición relativa que ocupa cada residuo en la región TM (Hessa et al., 2007). Un valor negativo de ΔG_{app} indica que la secuencia será reconocida por la compleja maquinaria del translocón como un fragmento TM, por el contrario, un valor positivo no implica necesariamente que la secuencia no sea TM, sino que el segmento no es capaz de insertarse eficientemente por si mismo, pudiendo requerir para su inserción la interacción con alguna otra hélice TM (http://www.cbr.su.se/DGpred). En este caso, las predicciones se basan pues en una escala de propensión obtenida experimentalmente empleando un sistema de traducción *in vitro* (Hessa et al., 2005; 2007), siendo ésta la primera escala de propensión biológica. Dado que la aproximación experimental usada mayoritariamente en esta tesis es la misma que la que se utilizó para el desarrollo matemático de este algoritmo, las predicciones realizadas en la tesis ha sido preferentemente obtenidas en este servidor de acceso libre en Internet (http://dgpred.cbr.su.se)

No existen muchos trabajos que evalúen de manera independiente la precisión de los diferentes métodos de predicción de segmentos TM. Teniendo en cuenta el conjunto de trabajos realizados en este sentido hasta la fecha, no parece que haya ningún método de predicción que sea significativamente mejor que el resto. Generalmente los métodos más sencillos, aquellos basados en escalas clásicas de hidrofobicidad, son los menos precisos. Aún así todos los programas parecen presentar dos grandes carencias. (1) En proteínas que presentan más de cinco regiones hidrofóbicas los fragmentos TM son identificados con poca precisión (Cuthbertson et al., 2006), probablemente por la falta de información acerca de las interacciones hélice-hélice que posiblemente se están produciendo. (2) Ninguno de los métodos está diseñado para la identificación de dominios de membrana helicoidales mas allá de la visión clásica de la hélice TM (ej. hélices cortas o lazos re-entrantes). Otro error común pero subsanado en alguno de los métodos más modernos como SPOCTOPUS (http://octopus.cbr.su.se/) es la incapacidad de los algoritmos para diferenciar entre una secuencia señal y un verdadero fragmento TM.

Es difícil conocer cómo se comportan estos programas (en cuanto a fiabilidad de los resultados obtenidos) cuando se enfrentan a secuencias frente a las que no han sido entrenados, dado que todas las comparativas realizadas a día de hoy entre los diferentes métodos de predicción emplean para probar los programas el mismo conjunto de proteínas utilizadas en el diseño de los algoritmos.

r.3.2. Mapeo por glicosilación

El método experimental empleado en el presente trabajo para la identificación de las regiones hidrofóbicas con capacidad para insertarse en la membrana fue principalmente la técnica de mapeo por glicosilación, utilizando la proteína modelo de *E. coli*, Lep *(Leader Peptidase)*. Este método, desarrollado en el laboratorio del Profesor Gunnar von Heijne (Universidad de Estocolmo) en la década de los 90, se basa en que la adición enzimática de oligosacáridos catalizada por la oligosacaril transferasa (OST) se puede dar únicamente en aquellas dianas de glicosilación que se encuentren en secuencias polipeptídicas translocadas al lumen del ER o en los dominios luminales de las proteínas de membrana. Así pues, debido a la localización del sitio activo de la OST (proteína de membrana asociada al translocón) aquellas dianas de glicosilación presentes en proteínas citosólicas o en dominios de proteínas de membrana oritentados hacia el citosol celular no serán modificadas.

La peptidasa de la secuencia señal de *E.coli* (Lep) es una proteína integral de membrana, localizada en la membrana interna de la bacteria. Presenta dos dominios TM, (H1 y H2) unidos por un lazo citosólico (P1) y un dominio periplásmico (P2).

La topología de la proteína posiciona ambos extremos, amino- y carboxilo-terminal, en el periplasma bacteriano. Es importante comentar para la comprensiónde la presente Tesis que la proteína adopta la misma topología cuando se sintetiza en presencia de microsomas de origen eucariótico (Johansson et al., 1993). Mediante mutagénesis dirigida se introdujo una diana de glicosilación (Asn-Ser-Thr, G1) al inicio del dominio P2, que por tanto será reconocida por la OST cuando la proteína sea traducida en presencia de membranas microsomales.

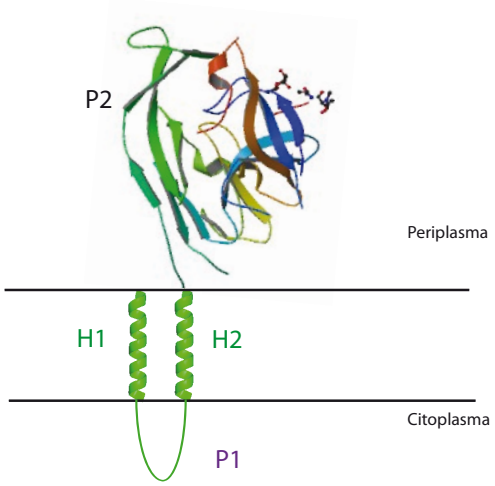


Figure 27 Esquema representativo de la disposició en membrana de la proteína modelo '*Leader peptidase*' (Lep) de *E.coli*. Los fragmentos TM de Lep (H1 y H2) se muestran como estructuras helicoidales y la estructura del dominio periplásmico P2 se ha tomado de las coordenadas depositadas en el PDB (código PDB: 3S04)

La secuencia concreta objeto de estudio es introducida en el dominio P2, y seguida de una segunda diana de glicosilación, quedando flanqueada por tanto por dos dianas de glicosilación artificiales (G1 y G2). En estas construcciones, si la región objeto de estudio se inserta en la membrana sólo G1 permanecerá en el lumen y será por tanto glicosilada (Figura 27A). Cuando el translocón no identifique a la secuencia problema como dominio TM el polipéptido será trasnslocado y ambas dianas G1 y G2 podrán ser modificadas (**Figura 28**B). El estado de glicosilación de la proteína lo podremos conocer a través de su movilidad electroforética. Para conocer la capacidad de inserción de las secuencias objeto de estudio las quimeras son traducidas *in vitro* utilizando lisado de reticulocito (el cual aporta la maquinaria de traducción), microsomas derivados de retículo endoplasmático (como fuente de membranas y de la maquinaria de inserción) y aminoácidos marcados radiactivamente con [^{35}S]. Posteriormente, las muestras se someten a una electroforesis desnaturalizante en geles de poliacrilamida en presencia de (SDS-PAGE). Las versiones de Lep glicosiladas y no glicosiladas presentaran una movilidad electroforética diferente debido a que la glicosilación en una diana de glicosilación aumenta el peso molecular de la proteína en aproximadamente 2.5 kDa (**Figura 28**).

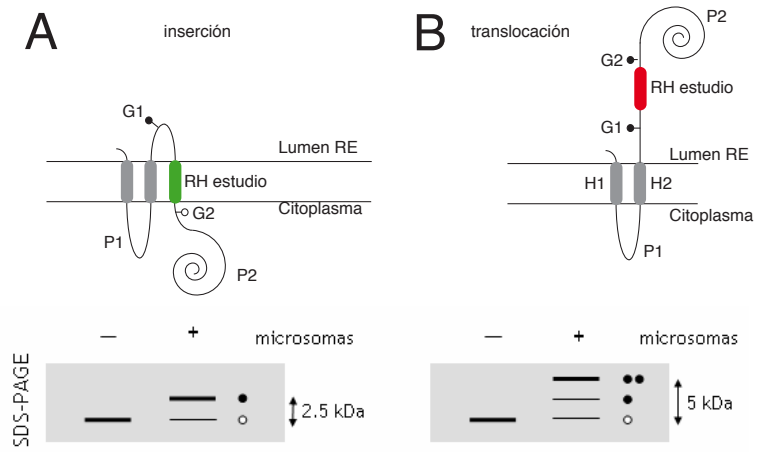


Figura 28. Identificación de fragmentos TM empleando el sistema de Lep. (A) La secuencia objeto de estudio se fusiona en el dominio P2 entre dos dianas de glicosilación (G1 y G2), de manera que si ésta es capaz de insertarse (en verde) en la membrana sólo G1 permanecerá en el lumen y será glicosilado. B. Si por el contrario la región a estudio no es reconocida como un dominio TM (en rojo) ambas dianas G1 y G2 podrán ser modificadas por la OST. Es posible diferenciar el estado de glicosilación de Lep por la movilidad electroforetica que presenta en geles de SDS-PAGE (sin glicosilar círculo vacío, mono-glicosilada un círculo negro, y doble-glicosilada dos círculos negros)

r.3.3. Determinación del grado de asociación a membranas

Un método sencillo y ampliamente empleado para determinar si una proteína se encuentra asociada a membranas es la sedimentación mediante ultracentrifugación. Las membranas microsomales y las proteínas asociadas a éstas (bien integralmente o bien de manera periférica son sedimentadas mediante ultracentrifugación a 100.000 g, mientras que las proteínas solubles citosólicas permanecen en el sobrenadante. En el sedimento de estas ultracentrifugaciones se encuentran también las proteínas solubles que son translocadas al interior de los microsomas por quedar "atrapadas" en el lumen. Para diferenciar entre estas situaciones será necesario realizar un tratamiento previo a la ultracentrifugación que pueden consistir en:

- **Extracción Alcalina**. El tratamiento con Na_2CO_3 (pH 11.5) transforma los microsomas en bicapas abiertas liberando su contenido. Tras la ultracentrifugación en el sedimento encontraremos ahora únicamente las proteínas asociadas a las membranas, mientras que las proteínas translocadas al lumen se acumularán en la fracción soluble.

- **Tratamiento con urea**. La urea es un potente agente desnaturalizante que al perturbar su estructura secundaria y terciaria solubiliza las proteínas que se encuentran perifericamente unidas a la membrana. Las proteínas integrales de membrana no pierden su asociación a los microsomas con este tratamiento dado que la membrana protege las regiones con estructura secundaria, permaneciendo en la fracción sedimentada tras el proceso de ultracentrifugación.
- **Reparto de fases con Triton X-114**. El triton X-114 es un detergente que por encima de su concentración micelar crítica presenta una elevada capacidad para solubilizar membranas cuando se encuentra entre 0 y 20ºC. A medida que la temperatura sube, aumenta el tamaño micelar del detergente hasta que, por encima de 20ºC se produce una clara separación de fases. Por encima de esta temperatura con una centrifugación a baja velocidad podremos separar una fase acuosa y otra orgánica. Mediante esta separación de fases, las proteínas integrales de membrana así como los lípidos se localizan fundamentalmente en la fase orgánica mientras que las proteínas periféricas y solubles se recuperan de la fase acuosa (Bordier, 1981).

r.3.4. Determinación de la topología de una proteína de membrana

La topología de una proteína de membrana queda definida por la localización de sus extremos amino- (N-t) y carboxilo-terminal (C-t) y por el número de dominios TM que presenta. En el presente trabajo con el propósito de determinar la localización de los extremos N-t y C-t se emplearon diferentes métodos *in vitro* e *in vivo*, basados en glicosilación y protección frente a la digestión con proteasas.

3.4.1. Determinación de la topología mediante modificación por glicosilación

La oligosacariltransferasa (OST) es una proteína de membrana asociada al complejo del translocón. Como se ha mencionado anteriormente presenta su sitio activo en el lado luminal de la membrana del ER, por lo que sólo es capaz de transferir azúcares a aquellas dianas de glicosilación de las proteínas de membrana que se encuentren en dominios

luminales, siempre y cuando éstas se encuentren a una distancia mínima de la membrana del ER de entre 30-40 Å, lo que supone en términos de número de residuos en el polipéptido entre 12 y 14 aminoácidos (Nilsson and Heijne, 1993). La ubicación de una diana de glicosilación (Asn-X-Thr, siendo X cualquier aminoácido excepto Pro) permite averiguar por tanto si el dominio hidrofílico donde se encuentra la diana está expuesto al lumen del retículo o al citosol. El sitio aceptor de glicosilación empleado podrá ser natural o introducido artificialmente, lo que permite determinar la localización de cualquier región no TM de la proteína. La sencillez de esta técnica admite que pueda ser aplicada tanto *in vitro* (Vilar et al., 2002) como *in vivo* (Tamborero et al., 2011). Existen diversos controles que pueden realizarse para asegurar que el aumento en el peso molecular de la proteína se debe a la adición de azúcares. (1) Traducción en ausencia de membranas (en aquellos casos en los que el experimento lo permita). (2) Tratamiento con **endoglicosidasa H**, enzima altamente específico que elimina oligosacáridos ricos en manosa unidos a residuos de asparagina. (3) Empleo de péptidos aceptores en exceso que compiten por la unión al sitio activo de la OST, disminuyendo así la eficiencia de glicosilación. (4) Modificación del aceptor de la *N*-Glicosilación (asparragina) por mutagénsis dirigida de la asparragina aceptora. En general se ha utilizado en la presente tesis la mutación a glutamina, por la similitud de características fisicoquímicas en las cadenas laterales de estos dos aminoácidos.

Para la determinación del número de segmentos que atraviesan la membrana presentes en la proteína y la topología de los mismos se diseñó una estrategia basada en la adición de etiquetas de glicosilación en el extremo C-t de polipéptidos derivados de la secuencia de la proteína de la cual se pretende conocer su topología y número de fragmentos TM. Mediante PCR y añadiendo en el oligonucleótido reverso una secuencia de glicosilación (NST), más tres aminoácidos (MMS), seguidos de dos codones de parada (NST+MMS+STOPSTOP), se generaron fragmentos de DNA conteniendo sucesivamente truncados de tamaño creciente de la proteína objeto de estudio. Los diferentes polipéptidos abarcaron progresivamente las diferentes regiones transmembranales putativas seguidas de una diana de glicosilación en el extremo C-t, de manera que la presencia o no de glicosilación en esta diana adicional es indicativa de la orientación que tiene el segmento

TM que la precede, ya que, como se ha mencionado anteriormente, los lazos que se localicen en el lumen del RE podrán ser glicosilados mientras que los que estén orientados al citosol no. La existencia o no de moléculas glicosiladas nos va indicando hacia qué parte de la membrana queda orientado el C-t del polipéptido analizado (**Figura 29**).

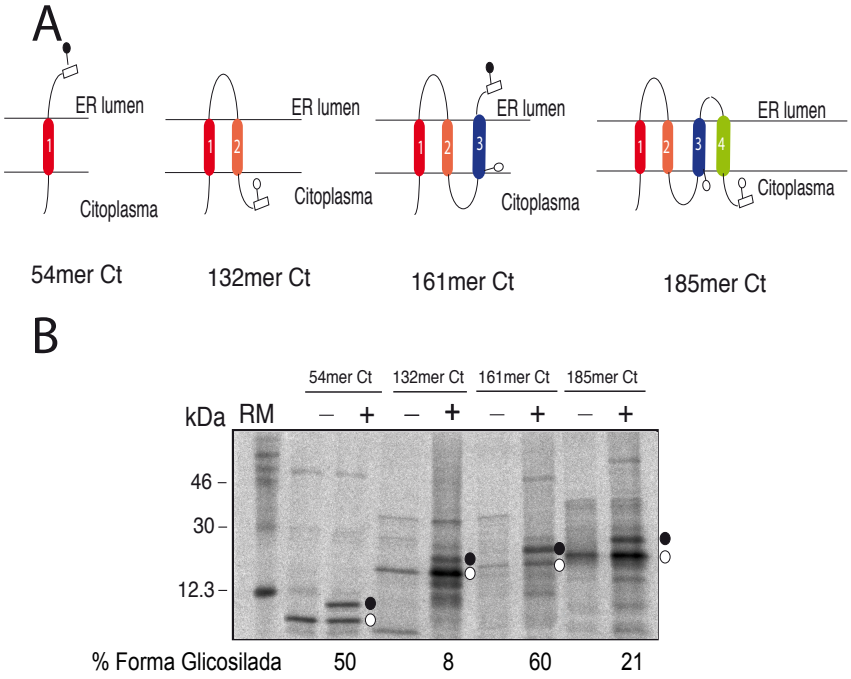


Figura 29 Eficiencia de la *N*-Glicosilación de truncados de la proteína TRAP-γ. La traducción *in vitro* de mRNAs con oligos reversos que introducen dianas de glicosilación en el extremo C-terminal (rectángulos blancos), fue llevada a cabo en presencia (+) y en ausencia (-) de membranas. Las bandas no glicosiladas son indicadas con un punto blanco, y las bandas mono-glicosiladas son indicadas con un punto negro, respectivamente. Los esquemas (A) muestran modelos de la topología que adquieren estos truncados en la membrana del ER. (B) Traducción *in vitro* de polipéptidos derivados de la secuencia de TRAP-γ en los que la etiqueta de glicosilación en el extremo C-t proporciona una interpretación sencilla de su topología.

3.4.2. Determinación de la topología mediante digestión con proteasas

Las proteasas son enzimas implicadas en la hidrólisis de cadenas polipeptídicas. Algunas de ellas pueden digerir únicamente el aminoácido localizado en el extremo amino- o carboxilo-terminal (exopeptidasas) mientras que otras pueden realizar su función en cualquier posición de la proteína (endopeptidasas) siempre y cuando se cumplan los requerimientos a

nivel de secuencia de aminoácidos necesarios. Las enzimas proteolíticas (especialmente las endopeptidasas) se ha consolidado como una técnica ampliamente utilizada para la determinación de la topología de proteínas de membrana (Heijne, 1989). Generalmente se emplea Proteinasa K (PK) o Tripsina aunque existen numerosos enzimas que pueden ser utilizadas, como pepsina, papaina o quimiotripsina (Franke et al, 2007). Muchas de estas proteínas mantienen su actividad en presencia de distintas sustancias desnaturalizantes como: SDS, urea o agentes quelantes (EDTA); además trabajan en un amplio rango de pH. Esta versatilidad las convierte en una potente herramienta de trabajo. El fundamento es similar para todas ellas. Estas proteasas son capaces de hidrolizar otras proteínas pero ninguna de ellas es capaz de atravesar la bicapa lipídica, de manera que, únicamente aquellos dominios de las proteínas de membrana accesibles a las proteasas serán digeridos. Analizando mediante electroforesis el número y tamaño de los péptidos que se generan tras la digestión, es decir, aquellos que son inaccesibles para las proteasas, (siempre y cuando estos fragmentos producidos sean de un tamaño suficientemente grande) podremos conocer la topología de la proteína.

r.3.5. Detección de proteínas mediante fotoentrecruzamiento

Para identificar las proteínas que interaccionan con cadenas nacientes de proteínas de membrana en los estadíos iniciales de su biosíntesis se utilizó en esta tesis la técnica de fotoentrecruzamieto molecular específico de posición (Krieg et al., 1986; McCormick et al., 2003; Saurí et al., 2005; 2009). En estos experimentos, RNAs mensajeros truncados con un codón ámbar (TAG) situado en la región objeto de estudio, en general en mitad de la primera región hidrofóbica de la proteína, fueron traducidos *in vitro* en presencia de aminoacil-tRNAs supresores del codón ámbar (εANB-LystRNAamb). En estas condiciones el ribosoma traduccionalmente activo se detiene al llegar al final de un mRNA truncado, pero la cadena naciente no se disocia, dado que el ribosoma no detecta el final de la traducción debido a la ausencia de codón de parada, de manera que se genera un verdadero intermediario del proceso de traducción. En casos en los que se estudió el direccionamiento de los complejos ribosoma-cadena naciente (RNC) en los ensayos no se añadieron membranas para facilitar que el complejo RCN dispusiera de mayor tiempo para

interaccionar con la SRP, favoreciendo el fotoentrecruzamiento con la misma. Los aminoacil-tRNAs supresores del codón ámbar utilizados (preparados en el laboratorio del Profesor Arthur E. Johnson, Texas A&M University) están cargados con un residuo de lisina derivatizado con la sonda 5-acido-2-nitrobenzoil (ANB) fotoactivable (εANB-Lys-tRNA^{amb}, figura 30), que al ser irradiada con luz ultravioleta entrecruza la cadena naciente en la que se ha incorporado con cualquier polipéptido que se encuentre a una distancia inferior a 5 Å de la sonda. En estas codiciones, los productos de traducción se fotolizaron para que la sonda pudiera reaccionar con aquellas moléculas que encuentre a su alrededor (**Figura 30**). A modo de ejemplo, en la **Figura 30** se muestran los resultados experimentales del entrecruzamiento de la proteína TRAP-γ con la SRP. Para posicionar el codón ámbar en el medio del primer segmento TM de TRAP-γ se realizó la predicción de su localización, utilizando el servidor *ΔG prediction server* (http://dgpred.cbr.su.se).

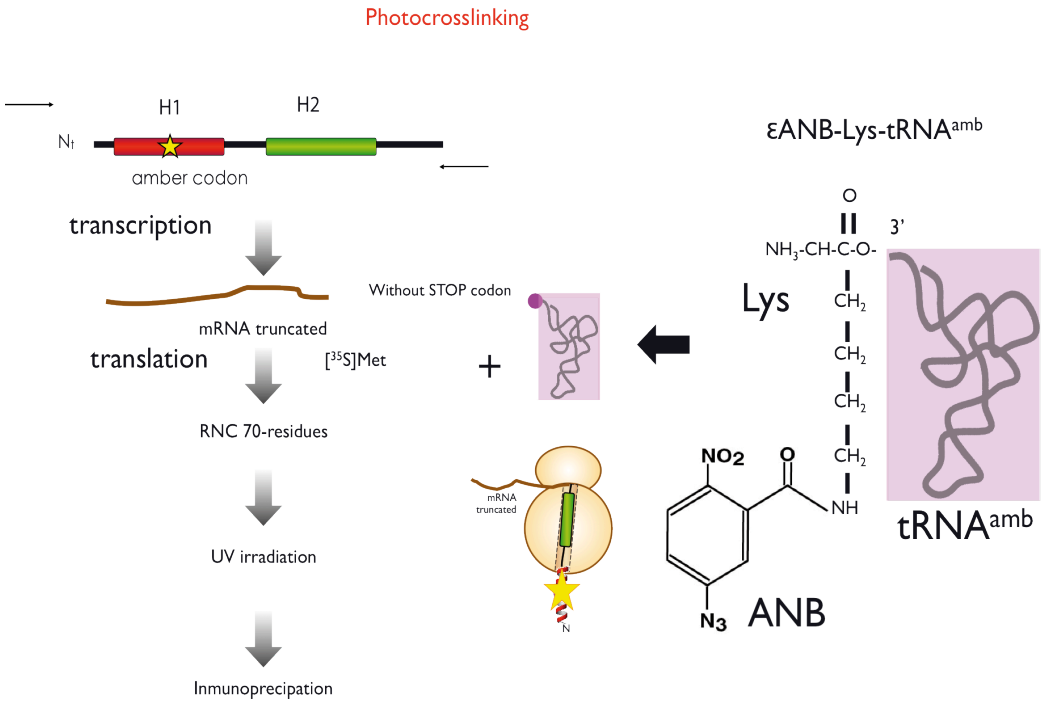


Figura 30. Fotoentrecruzamiento mediante luz UV. Fotoentrecreuzamiento con proteína TRAP-γ , como ejemplo, con la SRP en los primeros estadios de su biogénesis. Una única sonda fotoreactiva (arriba, izquierda) se incorporó mediante la introducción de un codón de parada ámbar en la posición central del fragment transmembrana; RCNs radiactivas se tradujeron en presencia de Lys-tRNA^{amb} sin modificar o de la sonda fotoreactiva εANB-Lys-tRNA^{amb}.

Previamente, se demostró experimentalmente que una longitud de 55 residuos entre la sonda fotoreactiva y el sitio P es suficiente para que la sonda quede expuesta fuera del túnel ribosomal y accesible a la SRP (Saurí et al., 2005). En la figura 5, se observa un fotoaducto de aproximadamente 67 kDa a niveles significativos cuando los complejos ribosoma/cadena naciente (RCN) son irradiados con luz ultravioleta, sólo cuando a la mezcla de reacción se le añade SRP y εANB-Lys-tRNA^{amb} . La masa molecular aparente de este fotoaducto corresponde a la de la cadena naciente de aminoácidos más la subunidad de 54 kDa de la SRP (SRP54). En los casos en que la sonda fotorreactiva no esta presente o es sustituida por el aminoacil-tRNA (Lys-tRNA^{amb}) sin modificar no se detectaran aductos. A continuación, se procede a identificación de los aductos por inmunoprecipitación con los anticuerpos correspondientes, en este caso anti-SRP54..

Para confirmar que la SRP es requerida en el direccionamiento de las proteínas a la membrana realizamos experimentos de traducción *in vitro* suplementando la mezcla con SRP purificada. El material biológico que aporta la maquinaria de traducción en los experimentos de fotoentrecruzamiento son de germen del trigo y la principal característica que le diferencia del lisado de reticulocitos (utilizado en los demás ensayos de esta tesis doctoral) es que su contenido en SRP es significativamente menor (Kanneret al., 2002; Tamborero et al., 2011).

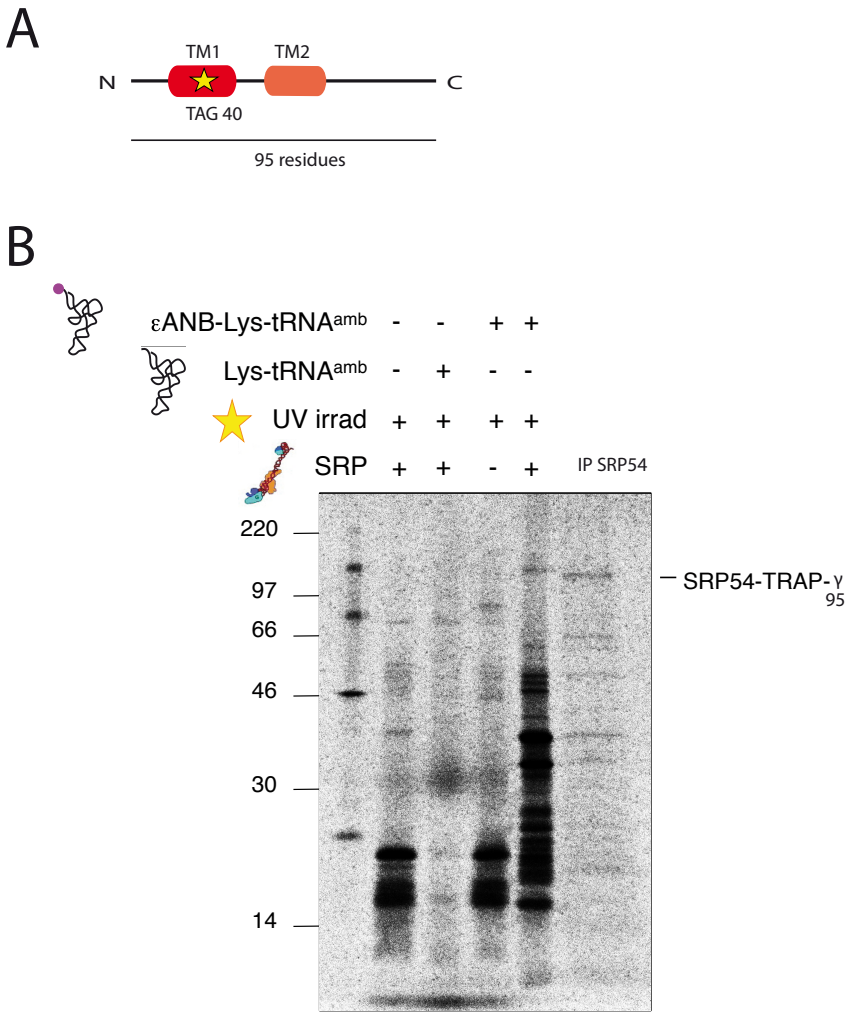


Figura 31. Foto-entrecruzamiento entre TRAP-γ y la SRP (A) Esquema de la construcción de TRAP-γ utilizada para los experimentos de foto-entrecruzamiento. Con la estrella se simboliza la posición de la sonda fotoactivable (TAG40). Las flechas indican los oligonucleótidos utilizados para la obtención del truncado de 95 aminoácidos. (B) Fotoentrecreuzamiento de TRAP-γ con la SRP en los primeros estadios de su biogénesis. Una sonda fotoreactiva se incorporó mediante la introducción de un codón de parada ámbar en la posición 40; RCNs radiactivas de 95 residuos se tradujeron en presencia de Lys-tRNA^{amb} sin modificar o de la sonda fotoreactiva εANB-Lys-tRNA^{amb}.

R.4. CONCLUSIONES

Capítulo 1

"N-Glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon"

- La eficiencia de glicosilación depende de la distancia entre el residuo de Asn aceptor y el extremo C-terminal del polipéptido, aumentando gradualmente con la distancia entre ambos.
- Se ha demostrado que las etiquetas de glicosilación en C-terminal necesitan un mínimo de seis residuos para una glicosilación eficiente.
- La naturaleza del aminoácido que precede la Asn receptora afecta a la eficiencia de glicosilación. En concreto, los residuos de Met, Trp y Arg disminuyen significativamente la eficiencia de glicosilación.
- Estadísticamente, los 20 aminoácidos naturales se han encontrado precediendo al residuo de Asn aceptor, pese a que hay diferencias significativas en las probabilidades de encontrar cada residuo.
- Se validó la utilidad de las etiquetas de glicosilación para estudios de la topología de proteínas de membrana, aportando un rápido y eficiente método para su determinación.

Capítulo 2

"Membrane integration of Poliovirus 2B Viroporin; a naturally occurring α-helical hairpin"

- Las regiones hidrofóbicas de la viroporina 2B en el contexto de la proteína modelo Lep no se integran como segmentos TM en presencia de membranas de derivadas del ER, pero sí lo hacen conjuntamente como una horquilla α-helicoidal.

- La viroporina 2B es integrada co-traduccionalmente en membrana de ER a través del translocon como una proteína de membrana con dos segmentos TM y una orientación N-/C- terminal citoplasmática.
- Las regiones hidrofóbicas 1 y 2 (RH1 y RH2) interaccionan entre si para formar el motivo hélice-giro-hélice que cruza la membrana. Se ha encontrado una posible estabilización adicional del motivo por los residuos localizados en el giro que conecta ambas hélices.
- La topología encontrada para la viroporina 2B fue demostrada en cultivos celulares, donde se mantuvó su capacidad permeabilizante.
- El residuo cargado negativamente (Asp74) es crítico para la integración de la horquilla.
- Los modelos moleculares obtenidos usando el protocolo "Rosseta" claramente resaltaron la preferencia de las Lys46 y Lys42 para interaccionar con el residuo Asp74.
- El plegamiento nativo de la proteína 2B puede estar favorecido por interacciones electroestáticas entre Lys-Asp localizadas en los fragmentos TM adyacentes.
- Los resultados sugieren que la formación de la α-helice horquilla se produce en los estadíos iniciales de la biogénesis de la proteína, en concreto en el translocon, permitiendo las interacciones polares entre TM consecutivos, estabilizando estos residuos en el polipéptido naciente.

Capítulo 3

"Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing"

- Se analizó la distribución de los residuos ionizables (Asp, Glu, Lys y Arg) en los segmentos TM a partir de estructuras de alta resolución de proteínas de membrana, estando presente en una frecuencia baja.
- A partir de nuestros experimentos en micelas de SDS, se concluye que las sustituciones de un residuo no polar por un residuo ionizable en la interfase de interacción de las hélices compromete severamente la formación del dímero. Por el

contrario, si el motivo de dimerización (la secuencia GxxxG) no es afectado y la sustitución se produce alejada de la interfase de interacción, dichas sustituciones no impiden la dimerización.

- Un puente salino intra-helice o interacciones mediante puentes de hidrogeno entre Lys y Asp localizados en la mismo lado de la hélice del segmento TM pueden facilitar la inserción del TM en membranas biológicas debido a la reducción del coste de la energia libre que requiere su reparto entre la membrana y la fase acuosa.
- Nuestros resultados indican que la presencia de residuos ionizables no impiden la inserción en la membrana y permiten la formación de dímeros tanto *in vitro* como en membranas de células bacterianas.

Capítulo 4

"Human Peroxisomal Membrane Protein PEX3: SRP-dependent Co-translational Integration into and Budding Vesicle Exit from the ER Membrane"

- La primera región hidrofóbica de PEX3 (RH1) es reconocida y unida por la SRP cuando emerge del ribosoma, indicando que actúa como secuencia señal.
- RH1, pero no RH2 de PEX3, es integrada eficientemente en la membrana del ER. Además, la RH1 es necesaria y suficiente para la integración estable de PEX3 en la membrana del ER.
- Los foto-aductos covalentes entre Sec61α y cadenas nacientes de PEX3 son sólo observables en presencia de SRP. Además, la SRP es requerida para dirigir la cadena naciente al translocon.
- La RH1 interacciona secuencialmente con Sec61α y TRAM durante su biogénesis. Estos resultados demuestran que PEX3 se inserta co-traduccionalmente en la membrana del ER mediante la ruta dependiente de SRP y del translocón.
- Hemos demostrado en este capítulo que una proteína peroxisomal humana (PEX3) es selectivamente extraída de la membrana del ER mediante la formación vesículas en un proceso dependiente de ATP. Este trabajo define el papel del ER en la biogénesis de los peroxisomas desde los estadios iniciales hasta el final.

REFERENCES

Adamian, L., and Liang, J. (2001). Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. Journal of Molecular Biology *311*, 891–907.

Agne, B., Meindl, N.M., Niederhoff, K., Einwächter, H., Rehling, P., Sickmann, A., Meyer, H.E., Girzalsky, W., and Kunau, W.H. (2003). Pex8p: an intraperoxisomal organizer of the peroxisomal import machinery. Mol. Cell *11*, 635–646.

Agrawal, G., Joshi, S., and Subramani, S. (2011). Cell-free sorting of peroxisomal membrane proteins from the endoplasmic reticulum. Proceedings of the National Academy of Sciences *108*, 9113–9118.

Aldabe, R., Barco, A., and Carrasco, L. (1996). Membrane permeabilization by poliovirus proteins 2B and 2BC. J. Biol. Chem. *271*, 23134–23137.

Aldabe, R., Irurzun, A., and Carrasco, L. (1997). Poliovirus protein 2BC increases cytosolic free calcium concentrations. J Virol *71*, 6214–6217.

Andersson, H., Bakker, E., and Heijne, von, G. (1992). Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in Escherichia coli. J. Biol. Chem. *267*, 1491–1495.

Baeza-Delgado, C., Marti-Renom, M.A., and Mingarro, I. (2013). Structure-based statistical analysis of transmembrane helices. Eur Biophys J *42*, 199–207.

Bano Polo, M., Baeza-Delgado, C., Orzáez, M., Marti-Renom, M.A., Abad, C., and Mingarro, I. (2012a). Polar/Ionizable residues in transmembrane segments: effects on helix-helix packing. PLoS ONE *7*, e44263.

Bano Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J.L., Elofsson, A., and Mingarro, I. (2012b). Charge Pair Interactions in Transmembrane Helices and Turn Propensity of the Connecting Sequence Promote Helical Hairpin Insertion. Journal of Molecular Biology.

Barco, A., and Carrasco, L. (1995). A human virus protein, poliovirus protein 2BC, induces membrane proliferation and blocks the exocytic pathway in the yeast Saccharomyces cerevisiae. Embo J. *14*, 3349–3364.

Barth, P., Wallner, B., and Baker, D. (2009). Prediction of membrane protein structures with complex topologies using limited constraints. Proc. Natl. Acad. Sci. U.S.a. *106*, 1409–1414.

Becker, T., Bhushan, S., Jarasch, A., Armache, J.-P., 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.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., and

Frank, J. (1997). Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. Science *278*, 2123–2126.

Ben-Dor, S., Esterman, N., Rubin, E., and Sharon, N. (2004). Biases and complex patterns in the residues flanking protein N-glycosylation sites. Glycobiology *14*, 95–101.

Bendz, M., Skwark, M., Nilsson, D., Granholm, V., Cristobal, S., Käll, L., and Elofsson, A. (2013). Membrane protein shaving with thermolysin can be used to evaluate topology predictors. Proteomics *13*, 1467–1480.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., Heijne, von, G., and Elofsson, A. (2008). Prediction of membrane-protein topology from first principles. Proc. Natl. Acad. Sci. U.S.a. *105*, 7177–7181.

Bogdanov, M., Heacock, P.N., and Dowhan, W. (2002). A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. Embo J. *21*, 2107–2116.

Booth, P.J., and Curran, A.R. (1999). Membrane protein folding. Current Opinion in Structural Biology *9*, 115–121.

Buck, T.M., Wagner, J., Grund, S., and Skach, W.R. (2007). A novel tripartite motif involved in aquaporin topogenesis, monomer folding and tetramerization. Nat. Struct. Mol. Biol. *14*, 762–769.

Cannon, K.S., Or, E., Clemons, W.M., Shibata, Y., and Rapoport, T.A. (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.

Caputo, G.A., and London, E. (2003). Cumulative effects of amino acid substitutions and hydrophobic mismatch upon the transmembrane stability and conformation of hydrophobic alpha-helices. Biochemistry *42*, 3275–3285.

Chamberlain, A.K., Lee, Y., Kim, S., and Bowie, J.U. (2004). Snorkeling preferences foster an amino acid composition bias in transmembrane helices. Journal of Molecular Biology *339*, 471–479.

Chin, C.N., and Heijne, von, G. (2000). Charge pair interactions in a model transmembrane helix in the ER membrane. Journal of Molecular Biology *303*, 1–5.

Choma, C., Gratkowski, H., Lear, J.D., and DeGrado, W.F. (2000). Asparagine-mediated self-association of a model transmembrane helix. Nat. Struct. Biol. *7*, 161–166.

Cioffi, J.A., Allen, K.L., Lively, M.O., and Kemper, B. (1989). Parallel effects of signal peptide hydrophobic core modifications on co-translational translocation and post-translational cleavage by purified signal peptidase. J. Biol. Chem. *264*, 15052–15058.

Cross, B.C.S., and High, S. (2009). Dissecting the physiological role of selective

transmembrane-segment retention at the ER translocon. Journal of Cell Science *122*, 1768–1777.

Curran, A.R., Templer, R.H., and Booth, P.J. (1999). Modulation of folding and assembly of the membrane protein bacteriorhodopsin by intermolecular forces within the lipid bilayer. Biochemistry *38*, 9328–9336.

Cymer, F., Veerappan, A., and Schneider, D. (2012). Transmembrane helix-helix interactions are modulated by the sequence context and by lipid bilayer properties. Biochim. Biophys. Acta *1818*, 963–973.

de Jong, A.S. (2002). Determinants for Membrane Association and Permeabilization of the Coxsackievirus 2B Protein and the Identification of the Golgi Complex as the Target Organelle. Journal of Biological Chemistry *278*, 1012–1021.

de Jong, A.S., de Mattia, F., Van Dommelen, M.M., Lanke, K., Melchers, W.J.G., Willems, P.H.G.M., and van Kuppeveld, F.J.M. (2008). Functional analysis of picornavirus 2B proteins: effects on calcium homeostasis and intracellular protein trafficking. J Virol *82*, 3782–3790.

de Planque, M.R.R., and Killian, J.A. (2003). Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring. Mol. Membr. Biol. *20*, 271–284.

Dempski, R.E., and Imperiali, B. (2002). Oligosaccharyl transferase: gatekeeper to the secretory pathway. Curr Opin Chem Biol *6*, 844–850.

Distel, B., Erdmann, R., Gould, S.J., Blobel, G., Crane, D.I., Cregg, J.M., Dodt, G., Fujiki, Y., Goodman, J.M., Just, W.W., et al. (1996). A unified nomenclature for peroxisome biogenesis factors. J. Cell Biol. *135*, 1–3.

Do, H., Falcone, D., Lin, J., Andrews, D.W., and Johnson, A.E. (1996). The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell *85*, 369–378.

Doedens, J.R., and Kirkegaard, K. (1995). Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A. Embo J. *14*, 894–907.

Doura, A.K., Kobus, F.J., Dubrovsky, L., Hibbard, E., and Fleming, K.G. (2004). Sequence context modulates the stability of a GxxxG-mediated transmembrane helix-helix dimer. Journal of Molecular Biology *341*, 991–998.

Duong, M.T., Jaszewski, T.M., Fleming, K.G., and MacKenzie, K.R. (2007). Changes in apparent free energy of helix-helix dimerization in a biological membrane due to point mutations. Journal of Molecular Biology *371*, 422–434.

Dupuy, A.D., and Engelman, D.M. (2008). Protein area occupancy at the center of the red blood cell membrane. Proc. Natl. Acad. Sci. U.S.a. *105*, 2848–2852.

Egea, P.F., and Stroud, R.M. (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.

Eilers, M., Shekar, S.C., Shieh, T., Smith, S.O., and Fleming, P.J. (2000). Internal packing of helical membrane proteins. Proc. Natl. Acad. Sci. U.S.a. *97*, 5796–5801.

Engelman, D.M., and Steitz, T.A. (1981). The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. Cell *23*, 411–422.

Engelman, D.M. (2005). Membranes are more mosaic than fluid. Nat Cell Biol *438*, 578–580.

Enquist, K., Fransson, M., Boekel, C., Bengtsson, I., Geiger, K., Lang, L., Pettersson, A., Johansson, S., Heijne, von, G., and Nilsson, I. (2009). Membrane-integration characteristics of two ABC transporters, CFTR and P-glycoprotein. Journal of Molecular Biology *387*, 1153–1164.

Erlandson, K.J., Miller, S.B.M., Nam, Y., Osborne, A.R., Zimmer, J., and Rapoport, T.A. (2008). A role for the two-helix finger of the SecA ATPase in protein translocation. Nature *455*, 984–987.

Foeger, N., Glaser, W., and Skern, T. (2002). Recognition of eukaryotic initiation factor 4G isoforms by picornaviral proteinases. J. Biol. Chem. *277*, 44300–44309.

Fons, R.D., Bogert, B.A., and Hegde, R.S. (2003). Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. J. Cell Biol. *160*, 529–539.

Freites, J.A., Tobias, D.J., Heijne, von, G., and White, S.H. (2005). Interface connections of a transmembrane voltage sensor. Proc. Natl. Acad. Sci. U.S.a. *102*, 15059–15064.

Gavel, Y., and Heijne, von, G. (1990). Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. Protein Eng. *3*, 433–442.

Gerstein, M., and Chothia, C. (1999). Perspectives: signal transduction. Proteins in motion. Science *285*, 1682–1683.

Goder, V., and Spiess, M. (2001). Topogenesis of membrane proteins: determinants and dynamics. FEBS Letters *504*, 87–93.

Goder, V., Junne, T., and Spiess, M. (2004). Sec61p contributes to signal sequence orientation according to the positive-inside rule. Mol. Biol. Cell *15*, 1470–1478.

Görlich, D., and Rapoport, T.A. (1993). Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. Cell *75*, 615–630.

Görlich, D., Hartmann, E., Prehn, S., and Rapoport, T.A. (1992). A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature *357*, 47–52.

Granseth, E., Heijne, von, G., and Elofsson, A. (2005). A study of the membrane-water interface region of membrane proteins. Journal of Molecular Biology *346*, 377–385.

Gratkowski, H., Lear, J.D., and DeGrado, W.F. (2001). Polar side chains drive the association of model transmembrane peptides. Proc. Natl. Acad. Sci. U.S.a. *98*, 880–885.

Gurba, K.N., Hernandez, C.C., Hu, N., and Macdonald, R.L. (2012). GABRB3 mutation, G32R, associated with childhood absence epilepsy alters α1β3γ2L γ-aminobutyric acid type A (GABAA) receptor expression and channel gating. Journal of Biological Chemistry *287*, 12083–12097.

Hamilton, S.R., Yao, S.Y., Ingram, J.C., Hadden, D.A., Ritzel, M.W., Gallagher, M.P., Henderson, P.J., Cass, C.E., Young, J.D., and Baldwin, S.A. (2001). Subcellular distribution and membrane topology of the mammalian concentrative Na+-nucleoside cotransporter rCNT1. J. Biol. Chem. *276*, 27981–27988.

Hamman, B.D., Chen, J.C., Johnson, E.E., and Johnson, A.E. (1997). The aqueous pore through the translocon has a diameter of 40-60 A during cotranslational protein translocation at the ER membrane. Cell *89*, 535–544.

Hartmann, E., Gorlich, D., Kostka, S., Otto, A., Kraft, R., Knespel, S., Burger, E., Rapoport, T.A., and PREHN, S. (1993). A tetrameric complex of membrane proteins in the endoplasmic reticulum. Eur. J. Biochem. *214*, 375–381.

Hedin, L.E., Ojemalm, K., Bernsel, A., Hennerdal, A., Illergård, K., Enquist, K., Kauko, A., Cristobal, S., Heijne, von, G., Lerch-Bader, M., et al. (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. Journal of Molecular Biology *396*, 221–229.

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.

Heijne, von, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. Journal of Molecular Biology *225*, 487–494.

Heijne, von, G., and Gavel, Y. (1988). Topogenic signals in integral membrane proteins. Eur. J. Biochem. *174*, 671–678.

Heinrich, S.U., Mothes, W., Brunner, J., and Rapoport, T.A. (2000). The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. Cell *102*, 233–244.

Heinrich, S.U., and Rapoport, T.A. (2003). Cooperation of transmembrane segments during the integration of a double-spanning protein into the ER membrane. Embo J. *22*, 3654–3663.

Hermansson, M., Monné, M., and Heijne, von, G. (2001). Formation of helical hairpins during membrane protein integration into the endoplasmic reticulum membrane. Role of the N and C-terminal flanking regions. Journal of Molecular Biology *313*, 1171–1179.

Herrmann, J.R., Fuchs, A., Panitz, J.C., Eckert, T., Unterreitmeier, S., Frishman, D., and Langosch, D. (2010). Ionic interactions promote transmembrane helix-helix association depending on sequence context. Journal of Molecular Biology *396*, 452–461.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., and Heijne, von, G. (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature *433*, 377–381.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., and Heijne, von, G. (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature *450*, 1026–1030.

Hettema, E.H., Girzalsky, W., van Den Berg, M., Erdmann, R., and Distel, B. (2000). Saccharomyces cerevisiae pex3p and pex19p are required for proper localization and stability of peroxisomal membrane proteins. Embo J. *19*, 223–233.

Higy, M., Junne, T., and Spiess, M. (2004). Topogenesis of Membrane Proteins at the Endoplasmic Reticulum †. Biochemistry *43*, 12716–12722.

Hildebrand, P.W., Rother, K., Goede, A., Preissner, R., and Frömmel, C. (2005). Molecular packing and packing defects in helical membrane proteins. Biophysj *88*, 1970–1977.

Hoepfner, D., Schildknegt, D., Braakman, I., Philippsen, P., and Tabak, H.F. (2005). Contribution of the endoplasmic reticulum to peroxisome formation. Cell *122*, 85–95.

Igura, M., and Kohda, D. (2011). Selective control of oligosaccharide transfer efficiency for the N-glycosylation sequon by a point mutation in oligosaccharyltransferase. Journal of Biological Chemistry *286*, 13255–13260.

Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita, T., Maenaka, K., and Kohda, D. (2008). Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. Embo J. *27*, 234–243.

Illergård, K., Kauko, A., and Elofsson, A. (2011). Why are polar residues within the membrane core evolutionary conserved? Proteins *79*, 79–91.

Ismail, N., Crawshaw, S.G., Cross, B.C.S., Haagsma, A.C., and High, S. (2008). Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. Biochem. J. *411*, 495–506.

Jacobs, R.E., and White, S.H. (1989). The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices. Biochemistry *28*, 3421–3437.

Jacobson, K., Mouritsen, O.G., and Anderson, R.G.W. (2007). Lipid rafts: at a crossroad between cell biology and physics. Nat Cell Biol *9*, 7–14.

Jansson, M., Wårell, K., Levander, F., and James, P. (2008). Membrane protein identification: N-terminal labeling of nontryptic membrane protein peptides facilitates database searching. J. Proteome Res. *7*, 659–665.

Jaud, S., Fernández-Vidal, M., Nilsson, I., Meindl-Beinker, N.M., Hübner, N.C., Tobias, D.J., Heijne, von, G., and White, S.H. (2009). Insertion of short transmembrane helices by the Sec61 translocon. Proc. Natl. Acad. Sci. U.S.a. *106*, 11588–11593.

Jensen, M.Ø., and Mouritsen, O.G. (2004). Lipids do influence protein function-the hydrophobic matching hypothesis revisited. Biochim. Biophys. Acta *1666*, 205–226.

Jiang, Y., Cheng, Z., Mandon, E.C., and Gilmore, R. (2008). An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. J. Cell Biol. *180*, 1149–1161.

Johansson, A.C.V., and Lindahl, E. (2006). Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. Biophysj *91*, 4450–4463.

Johansson, M., Nilsson, I., and Heijne, von, G. (1993). Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes. Mol. Gen. Genet. *239*, 251–256.

Johnson, A.E., and van Waes, M.A. (1999). The translocon: a dynamic gateway at the ER membrane. Annu. Rev. Cell Dev. Biol. *15*, 799–842.

Junne, T., Schwede, T., Goder, V., and Spiess, M. (2007). Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. J. Biol. Chem. *282*, 33201–33209.

Kalies, K.U., Rapoport, T.A., and 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.

Kammerer, S., Holzinger, A., Welsch, U., and Roscher, A.A. (1998). Cloning and characterization of the gene encoding the human peroxisomal assembly protein Pex3p. FEBS Letters *429*, 53–60.

Kasturi, L., Chen, H., and Shakin-Eshleman, S.H. (1997). Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. *323 (Pt 2)*, 415–419.

Kauko, A., Hedin, L.E., Thebaud, E., Cristobal, S., Elofsson, A., and Heijne, von, G. (2010). Repositioning of transmembrane alpha-helices during membrane protein folding. Journal of Molecular Biology *397*, 190–201.

Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology *338*, 1027–1036.

Killian, J.A. (1998). Hydrophobic mismatch between proteins and lipids in membranes. Biochim. Biophys. Acta *1376*, 401–415.

Kim, S., Jeon, T.-J., Oberai, A., Yang, D., Schmidt, J.J., and Bowie, J.U. (2005). Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. Proc. Natl. Acad. Sci. U.S.a. *102*, 14278–14283.

Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., et al. (2006). Definition of the bacterial N-glycosylation site consensus sequence. Embo J. *25*, 1957–1966.

Kozak, M. (1989). Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. Mol. Cell. Biol. *9*, 5073–5080.

Kragt, A., Voorn-Brouwer, T., van den Berg, M., and Distel, B. (2005). Endoplasmic reticulum-directed Pex3p routes to peroxisomes and restores peroxisome formation in a Saccharomyces cerevisiae pex3Delta strain. J. Biol. Chem. *280*, 34350–34357.

Krieg, U.C., Walter, P., and Johnson, A.E. (1986). Photocrosslinking of the signal sequence of nascent preprolactin to the 54-kilodalton polypeptide of the signal recognition particle. Proc. Natl. Acad. Sci. U.S.a. *83*, 8604–8608.

Krogh, A., Larsson, B., Heijne, von, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology *305*, 567–580.

Lam, S.K., Yoda, N., and Schekman, R. (2011). A vesicle carrier that mediates peroxisome protein traffic from the endoplasmic reticulum. Proceedings of the National Academy of Sciences *108*, E51–E52.

Lama, J., and Carrasco, L. (1992). Expression of poliovirus nonstructural proteins in Escherichia coli cells. Modification of membrane permeability induced by 2B and 3A. J. Biol. Chem. *267*, 15932–15937.

Landolt-Marticorena, C., and Reithmeier, R.A. (1994). Asparagine-linked oligosaccharides are localized to single extracytosolic segments in multi-span membrane glycoproteins. Biochem. J. *302 (Pt 1)*, 253–260.

Langosch, D., Brosig, B., Kolmar, H., and Fritz, H.J. (1996). Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. Journal of Molecular Biology *263*, 525–530.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., and Engelman, D.M. (1992). Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry *31*, 12719–12725.

Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I., and Heijne, von, G. (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. Proc. Natl. Acad. Sci. U.S.a. *105*, 4127–4132.

Levy, R., Wiedmann, M., and Kreibich, G. (2001). In vitro binding of ribosomes to the beta subunit of the Sec61p protein translocation complex. J. Biol. Chem. *276*, 2340–2346.

Li, E., Wimley, W.C., and Hristova, K. (2012). Transmembrane helix dimerization: Beyond the search for sequence motifs. Biochimica Et Biophysica Acta (BBA) - Biomembranes *1818*, 183–193.

MacKenzie, K.R., Prestegard, J.H., and Engelman, D.M. (1997). A transmembrane helix dimer: structure and implications. Science *276*, 131–133.

Madan, V., Sánchez-Martínez, S., Carrasco, L., and Nieva, J. (2010). A peptide based on the pore-forming domain of pro-apoptotic poliovirus 2B viroporin targets mitochondria. *1798*, 52–58.

Madan, V., Sánchez-Martínez, S., Vedovato, N., Rispoli, G., Carrasco, L., and Nieva, J.L. (2007). Plasma membrane-porating domain in poliovirus 2B protein. A short peptide mimics viroporin activity. Journal of Molecular Biology *374*, 951–964.

Marsh, D. (2006). Elastic curvature constants of lipid monolayers and bilayers. Chemistry and Physics of Lipids *144*, 146–159.

Martinez-Gil, L., Bano-Polo, M., Redondo, N., Sánchez-Martínez, S., Nieva, J.L., Carrasco, L., and Mingarro, I. (2011). Membrane Integration of Poliovirus 2B Viroporin. J Virol *85*, 11315–11324.

Martínez-Gil, L., Johnson, A.E., and Mingarro, I. (2010). Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. J Virol *84*, 5520–5527.

Martínez-Gil, L., Pérez-Gil, J., and Mingarro, I. (2008). The surfactant peptide KL4 sequence is inserted with a transmembrane orientation into the endoplasmic reticulum membrane. Biophys. J. *95*, L36–L38.

Martínez-Gil, L., Saurí, A., Marti-Renom, M.A., and Mingarro, I. (2011). Membrane protein integration into the endoplasmic reticulum. FEBS Journal 278, 3840-3856.

Martoglio, B., Hofmann, M.W., Brunner, J., and 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.

Matsuzaki, T., and Fujiki, Y. (2008). The peroxisomal membrane protein import receptor Pex3p is directly transported to peroxisomes by a novel Pex19p- and Pex16p-dependent pathway. J. Cell Biol. *183*, 1275–1286.

McCormick, P.J., Miao, Y., Shao, Y., Lin, J., and Johnson, A.E. (2003). Cotranslational

protein integration into the ER membrane is mediated by the binding of nascent chains to translocon proteins. Mol. Cell *12*, 329–341.

Meacock, S.L., Lecomte, F.J.L., Crawshaw, S.G., and 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.

Meindl-Beinker, N.M., Lundin, C., Nilsson, I., White, S.H., and Heijne, von, G. (2006). Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. EMBO Rep. *7*, 1111–1116.

Mellquist, J.L., Kasturi, L., Spitalnik, S.L., and Shakin-Eshleman, S.H. (1998). The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. Biochemistry *37*, 6833–6837.

Ménétret, J.-F., Hegde, R.S., Aguiar, M., Gygi, S.P., Park, E., Rapoport, T.A., and Akey, C.W. (2008). Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. Structure *16*, 1126–1137.

Ménétret, J.-F., Schaletzky, J., Clemons, W.M., Osborne, A.R., Skånland, S.S., Denison, C., Gygi, S.P., Kirkpatrick, D.S., Park, E., Ludtke, S.J., et al. (2007). Ribosome binding of a single copy of the SecY complex: implications for protein translocation. Mol. Cell *28*, 1083–1092.

Mingarro, I., Whitley, P., Lemmon, M.A., and Heijne, von, G. (1996). Ala-insertion scanning mutagenesis of the glycophorin A transmembrane helix: a rapid way to map helix-helix interactions in integral membrane proteins. Protein Sci. *5*, 1339–1341.

Monné, M., Hermansson, M., and Heijne, von, G. (1999). A turn propensity scale for transmembrane helices. Journal of Molecular Biology *288*, 141–145.

Monné, M., Nilsson, I., Johansson, M., Elmhed, N., and Heijne, von, G. (1998). Positively and negatively charged residues have different effects on the position in the membrane of a model transmembrane helix. Journal of Molecular Biology *284*, 1177–1183.

Mothes, W., Heinrich, S.U., Graf, R., Nilsson, I., Heijne, von, G., Brunner, J., and Rapoport, T.A. (1997). Molecular mechanism of membrane protein integration into the endoplasmic reticulum. Cell *89*, 523–533.

Mouritsen, O.G., and Bloom, M. (1993). Models of lipid-protein interactions in membranes. Annu. Rev. Biophys. Biomol. Struct. *22*, 145–171.

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C., and Jovine, L. (2003). Structure, function and evolution of the signal recognition particle. Embo J. *22*, 3479–3485.

Nagaoka, K., Hanioka, N., Ikushiro, S., Yamano, S., and Narimatsu, S. (2012). The effects of N-glycosylation on the glucuronidation of zidovudine and morphine by UGT2B7 expressed in HEK293 cells. Drug Metab. Pharmacokinet. *27*, 388–397.

Nagy, A., and Turner, R.J. (2007). The membrane integration of a naturally occurring alpha-helical hairpin. Biochem. Biophys. Res. Commun. *356*, 392–397.

Nelson, L.D., Johnson, A.E., and London, E. (2008). How interaction of perfringolysin O with membranes is controlled by sterol structure, lipid structure, and physiological low pH: insights into the origin of perfringolysin O-lipid raft interaction. J. Biol. Chem. *283*, 4632–4642.

Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J. Cell Biol. *134*, 269–278.

Nielsen, H., Engelbrecht, J., Heijne, von, G., and Brunak, S. (1996). Defining a similarity threshold for a functional protein sequence pattern: the signal peptide cleavage site. Proteins *24*, 165–177.

Nieva, J., Agirre, A., Nir, S., and Carrasco, L. (2003). Mechanisms of membrane permeabilization by picornavirus 2B viroporin. *552*, 68–73.

Nilsson, I., and Heijne, von, G. (2000). Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. J. Biol. Chem. *275*, 17338–17343.

Nilsson, I.M., and Heijne, von, G. (1993a). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J. Biol. Chem. *268*, 5798–5801.

Nilsson, I.M., and Heijne, von, G. (1993b). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. J. Biol. Chem. *268*, 5798–5801.

Novikoff, P.M., and Novikoff, A.B. (1972). Peroxisomes in absorptive cells of mammalian small intestine. J. Cell Biol. *53*, 532–560.

Nugent, T., and Jones, D.T. (2010). Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. PLoS Comput. Biol. *6*, e1000714.

Oberai, A., Joh, N.H., Pettit, F.K., and Bowie, J.U. (2009). Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. Proceedings of the National Academy of Sciences *106*, 17747–17750.

Ojemalm, K., Halling, K.K., Nilsson, I., and Heijne, von, G. (2012). Orientational Preferences of Neighboring Helices Can Drive ER Insertion of a Marginally Hydrophobic Transmembrane Helix. Mol. Cell *45*, 529–540.

Orzáez, M., Pérez-Payá, E., and Mingarro, I. (2000). Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. Protein Sci. *9*, 1246–1253.

Orzáez, M., Lukovic, D., Abad, C., Perez-Payá, E., and Mingarro, I. (2005). Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin A dimerisation motif. FEBS Letters *579*, 1633–1638.

Orzáez, M., Salgado, J., Giménez-Giner, A., Perez-Payá, E., and Mingarro, I. (2004). Influence of proline residues in transmembrane helix packing. Journal of Molecular Biology *335*, 631–640.

Ota, K., Sakaguchi, M., Heijne, von, G., Hamasaki, N., and Mihara, K. (1998). Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. Mol. Cell *2*, 495–503.

Park, E., and Rapoport, T.A. (2012). Bacterial protein translocation requires only one copy of the SecY complex in vivo. J. Cell Biol.

Petrescu, A.-J., Milac, A.-L., Petrescu, S.M., Dwek, R.A., and Wormald, M.R. (2004). Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology *14*, 103–114.

Pitonzo, D., Yang, Z., Matsumura, Y., Johnson, A.E., and Skach, W.R. (2009a). Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum Sec61 translocon. Mol. Biol. Cell *20*, 685–698.

Pitonzo, D., Yang, Z., Matsumura, Y., Johnson, A.E., and Skach, W.R. (2009b). Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum Sec61 translocon. Mol. Biol. Cell *20*, 685–698.

Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J., and Rapoport, T.A. (1998). Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell *94*, 795–807.

Pool, M.R. (2009). A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. J. Cell Biol. *185*, 889–902.

Popot, J.L., and Engelman, D.M. (1990). Membrane protein folding and oligomerization: the two-stage model. Biochemistry *29*, 4031–4037.

Popot, J.L., Gerchman, S.E., and Engelman, D.M. (1987). Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. Journal of Molecular Biology *198*, 655–676.

Psachoulia, E., Fowler, P.W., Bond, P.J., and Sansom, M.S.P. (2008). Helix−Helix Interactions in Membrane Proteins: Coarse-Grained Simulations of Glycophorin A Helix Dimerization †. Biochemistry *47*, 10503–10512.

Quinn, P., Griffiths, G., and Warren, G. (1984). Density of newly synthesized plasma membrane proteins in intracellular membranes II. Biochemical studies. J. Cell Biol. *98*, 2142–2147.

Ramadurai, S., Holt, A., Schäfer, L.V., Krasnikov, V.V., Rijkers, D.T.S., Marrink, S.J., Killian, J.A., and Poolman, B. (2010). Influence of hydrophobic mismatch and amino acid composition on the lateral diffusion of transmembrane peptides. Biophys. J. *99*, 1447–1454.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669.

Rapoport, T.A. (2008). Protein transport across the endoplasmic reticulum membrane. Febs J. *275*, 4471–4478.

Rapoport, T.A., Goder, V., Heinrich, S.U., and Matlack, K.E.S. (2004). Membrane-protein integration and the role of the translocation channel. Trends Cell Biol. *14*, 568–575.

Raue, U., Oellerer, S., and Rospert, S. (2007). Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. J. Biol. Chem. *282*, 7809–7816.

Roof, D.J., and Heuser, J.E. (1982). Surfaces of rod photoreceptor disk membranes: integral membrane components. J. Cell Biol. *95*, 487–500.

Rutz, C., Rosenthal, W., and Schülein, R. (1999). A single negatively charged residue affects the orientation of a membrane protein in the inner membrane of Escherichia coli only when it is located adjacent to a transmembrane domain. J. Biol. Chem. *274*, 33757–33763.

Sadlish, H., Pitonzo, D., Johnson, A.E., and Skach, W.R. (2005). Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispanning membrane protein. Nat. Struct. Mol. Biol. *12*, 870–878.

Sandoval, I.V., and Carrasco, L. (1997). Poliovirus infection and expression of the poliovirus protein 2B provoke the disassembly of the Golgi complex, the organelle target for the antipoliovirus drug Ro-090179. J Virol *71*, 4679–4693.

Saurí, A., McCormick, P.J., Johnson, A.E., and Mingarro, I. (2007). Sec61alpha and TRAM are sequentially adjacent to a nascent viral membrane protein during its ER integration. Journal of Molecular Biology *366*, 366–374.

Saurí, A., Saksena, S., Salgado, J., Johnson, A.E., and 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.

Saurí, A., Tamborero, S., Martínez-Gil, L., Johnson, A.E., and Mingarro, I. (2009). Viral membrane protein topology is dictated by multiple determinants in its sequence. Journal of Molecular Biology *387*, 113–128.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D.,

Schwappach, B., and Weissman, J.S. (2008). The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. Cell *134*, 634–645.

Shakin-Eshleman, S.H., Spitalnik, S.L., and Kasturi, L. (1996). The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. J. Biol. Chem. *271*, 6363–6366.

Shao, S., and Hegde, R.S. (2011). Membrane protein insertion at the endoplasmic reticulum. Annu. Rev. Cell Dev. Biol. *27*, 25–56.

Sharpe, H.J., Stevens, T.J., and Munro, S. (2010). A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. Cell *142*, 158–169.

Singer, S.J., and Nicolson, G.L. (1972). The fluid mosaic model of the structure of cell membranes. Science *175*, 720–731.

Skach, W.R. (2009). Cellular mechanisms of membrane protein folding. Nat. Struct. Mol. Biol. *16*, 606–612.

Stefansson, A., Armulik, A., Nilsson, I., Heijne, von, G., and Johansson, S. (2004). Determination of N- and C-terminal borders of the transmembrane domain of integrin subunits. J. Biol. Chem. *279*, 21200–21205.

Strandberg, E., and Killian, J.A. (2003). Snorkeling of lysine side chains in transmembrane helices: how easy can it get? FEBS Letters *544*, 69–73.

Subramani, S. (1998). Components involved in peroxisome import, biogenesis, proliferation, turnover, and movement. Physiol. Rev. *78*, 171–188.

Suhy, D.A., Giddings, T.H., and Kirkegaard, K. (2000). Remodeling the endoplasmic reticulum by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-induced vesicles. J Virol *74*, 8953–8965.

Sung, M., and Dalbey, R.E. (1992). Identification of potential active-site residues in the Escherichia coli leader peptidase. J. Biol. Chem. *267*, 13154–13159.

Tabak, H.F., Braakman, I., and van der Zand, A. (2012). Peroxisome Formation and Maintenance Are Dependent on the Endoplasmic Reticulum. Annu. Rev. Biochem. *82*, 2013 (in press) .

Tam, Y.Y.C., Fagarasanu, A., Fagarasanu, M., and Rachubinski, R.A. (2005). Pex3p initiates the formation of a preperoxisomal compartment from a subdomain of the endoplasmic reticulum in Saccharomyces cerevisiae. J. Biol. Chem. *280*, 34933–34939.

Tamborero, S., Vilar, M., Martínez-Gil, L., Johnson, A.E., and Mingarro, I. (2011). Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). Journal of Molecular Biology *406*, 571–582.

Thaysen-Andersen, M., and Packer, N.H. (2012). Site-specific glycoproteomics confirms that protein structure dictates formation of N-glycan type, core fucosylation and branching. Glycobiology *22*, 1440–1452.

Thoms, S., Harms, I., Kalies, K.-U., and Gärtner, J. (2012). Peroxisome formation requires the endoplasmic reticulum channel protein Sec61. Traffic *13*, 599–609.

Tsukazaki, T., Mori, H., Fukai, S., Ishitani, R., Mori, T., Dohmae, N., Perederina, A., Sugita, Y., Vassylyev, D.G., Ito, K., et al. (2008). Conformational transition of Sec machinery inferred from bacterial SecYE structures. Nature *455*, 988–991.

Tulumello, D.V., and Deber, C.M. (2009). SDS micelles as a membrane-mimetic environment for transmembrane segments. Biochemistry *48*, 12096–12103.

Tusnády, G.E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics *17*, 849–850.

van Dalen, A., and de Kruijff, B. (2004). The role of lipids in membrane insertion and translocation of bacterial proteins. Biochim. Biophys. Acta *1694*, 97–109.

van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36–44.

van der Zand, A., Braakman, I., and Tabak, H.F. (2010). Peroxisomal membrane proteins insert into the endoplasmic reticulum. Mol. Biol. Cell *21*, 2057–2065.

van der Zand, A., Gent, J., Braakman, I., and Tabak, H.F. (2012). Biochemically distinct vesicles from the endoplasmic reticulum fuse to form peroxisomes. Cell *149*, 397–409.

van Klompenburg, W., Nilsson, I., Heijne, von, G., and de Kruijff, B. (1997). Anionic phospholipids are determinants of membrane protein topology. Embo J. *16*, 4261–4266.

van Meer, G., Voelker, D.R., and Feigenson, G.W. (2008). Membrane lipids: where they are and how they behave. Nat. Rev. Mol. Cell Biol. *9*, 112–124.

Vereb, G., Szöllosi, J., Matkó, J., Nagy, P., Farkas, T., Vigh, L., Mátyus, L., Waldmann, T.A., and Damjanovich, S. (2003). Dynamic, yet structured: The cell membrane three decades after the Singer-Nicolson model. Proc. Natl. Acad. Sci. U.S.a. *100*, 8053–8058.

Viklund, H., Bernsel, A., Skwark, M., and Elofsson, A. (2008). SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics *24*, 2928–2929.

Voigt, S., Jungnickel, B., Hartmann, E., and Rapoport, T.A. (1996). Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. J. Cell Biol. *134*, 25–35.

Wang, B., Heath-Engel, H., Zhang, D., Nguyen, N., Thomas, D.Y., Hanrahan, J.W., and Shore, G.C. (2008). BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTRDeltaF508 via the derlin-1 complex. Cell *133*, 1080–1092.

Wehbi, H., Rath, A., Glibowicka, M., and Deber, C.M. (2007). Role of the extracellular loop in the folding of a CFTR transmembrane helical hairpin. Biochemistry *46*, 7099–7106.

White, S.H., and Wimley, W.C. (1999). MEMBRANE PROTEIN FOLDING AND STABILITY: Physical Principles. Annu. Rev. Biophys. Biomol. Struct. *28*, 319–365.

Wiener, M.C., and White, S.H. (1992). Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. Biophysj *61*, 434–447.

Wilkinson, B.M., Critchley, A.J., and Stirling, C.J. (1996). Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. J. Biol. Chem. *271*, 25590–25597.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. *3*, 842–848.

Wimley, W.C., Creamer, T.P., and White, S.H. (1996). Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. Biochemistry *35*, 5109–5124.

Wu, C.C., and Yates, J.R. (2003). The application of mass spectrometry to membrane proteomics. Nat. Biotechnol. *21*, 262–267.

Yahr, T.L., and Wickner, W.T. (2000). Evaluating the oligomeric state of SecYEG in preprotein translocase. Embo J. *19*, 4393–4401.

Yamaguchi, A., Hori, O., Stern, D.M., Hartmann, E., Ogawa, S., and 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. J. Cell Biol. *147*, 1195–1204.

Zhang, J., and Lazaridis, T. (2009). Transmembrane helix association affinity can be modulated by flanking and noninterfacial residues. Biophysj *96*, 4418–4427.

Zhang, L., Sato, Y., Hessa, T., Heijne, von, G., Lee, J.-K., Kodama, I., Sakaguchi, M., and Uozumi, N. (2007). Contribution of hydrophobic and electrostatic interactions to the membrane integration of the Shaker K+ channel voltage sensor domain. Proc. Natl. Acad. Sci. U.S.a. *104*, 8263–8268.

Zheng, N., and Gierasch, L.M. (1996). Signal sequences: the same yet different. Cell *86*, 849–852.

Zhou, F.X., Cocco, M.J., Russ, W.P., Brunger, A.T., and Engelman, D.M. (2000).

Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat. Struct. Biol. *7*, 154–160.

Zhou, F.X., Merianos, H.J., Brunger, A.T., and Engelman, D.M. (2001). Polar residues drive association of polyleucine transmembrane helices. Proc. Natl. Acad. Sci. U.S.a. *98*, 2250–2255.

Zimmer, J., Nam, Y., and Rapoport, T.A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature *455*, 936–943.