



Universitat de València

Departamento de Bioquímica y Biología Molecular  
Instituto Universitario de Biotecnología y Biomedicina (BIOTECMED)

Brayan Grau Argente  
TESIS DOCTORAL  
Enero 2021

UNDERSTANDING MEMBRANE  
PROTEIN FOLDING AND  
INTERFACIALITY

Directores: Dr. Ismael Mingarro  
Dr. Luis Martínez Gil

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

Programa de Doctorado en Biomedicina y Biotecnología



ISMAEL MINGARRO MUÑOZ, Doctor en Ciències Biològiques i Catedràtic del Departament de Bioquímica i Biologia Molecular de la Universitat de València.

LUIS MARTÍNEZ GIL, Doctor en Bioquímica i Professor Ajudant del Departament de Bioquímica i Biologia Molecular de la Universitat de València.

MANIFESTEN que En BRAYAN GRAU ARGENTE, Graduat en Biologia per la Universitat d'Alacant, ha realitzat sota la nostra direcció al Laboratori de Proteïnes de Membrana de l'Institut de Biotecnologia i Biomedicina (BIOTECMED) el treball que amb el títol: *Understanding membrane protein folding and interfaciality*, presenta per a optar al grau de Doctor per la Universitat de València. Alhora, aprofitem per a autoritzar la presentació i defensa pública d'aquest treball.

I perquè així conste i tinga els efectes oportuns, signem el present document en Burjassot l'12 de gener de 2021.

Dr. Ismael Mingarro

Dr. Luis Martínez-Gil



*Gracias a todas las personas que han estado ahí cuando les he necesitado.*



# Table of contents

<b>INTRODUCTION .....</b>	<b>9</b>
I.1. Biological membranes .....	11
<i>i.1.1. Structure of biological membranes.....</i>	<i>12</i>
<i>i.1.2. Membrane organization .....</i>	<i>14</i>
I.2. Membrane protein synthesis and targeting.....	16
<i>i.2.1. The ribosome .....</i>	<i>17</i>
<i>i.2.2. Targeting to the Sec translocon in eukaryotic cells .....</i>	<i>19</i>
<i>i.2.3. The Translocon .....</i>	<i>22</i>
I.3. Insertion of membrane proteins.....	32
<i>i.3.1. TMD hydrophobicity and amino acid preferences .....</i>	<i>33</i>
<i>i.3.2. Helical conformation of TMD .....</i>	<i>34</i>
<i>i.3.3. Hydrophobic matching .....</i>	<i>35</i>
I.4. Topology of membrane proteins.....	37
<i>i.4.1. Topology of single-spanning membrane proteins .....</i>	<i>37</i>
<i>i.4.2. Topology of multi-spanning membrane proteins .....</i>	<i>39</i>
<i>i.4.3. Orientation of transmembrane helices .....</i>	<i>40</i>
<i>i.4.4. Monotopic membrane proteins .....</i>	<i>42</i>
<i>i.4.5. Packing motifs of membrane proteins .....</i>	<i>43</i>
<b>OBJECTIVES .....</b>	<b>45</b>
<b>MATERIALS AND METHODS .....</b>	<b>49</b>
<i>In vitro</i> transcription and translation.....	51
<i>Plasmid preparation .....</i>	<i>51</i>
<i>In vitro</i> protein expression.....	51
<i>Endoglycosidase H and Proteinase K treatments .....</i>	<i>52</i>
Confocal microscopy .....	52
<i>Immunofluorescence .....</i>	<i>52</i>
Dimerization assays .....	53
<i>ToxRed assay .....</i>	<i>53</i>
<i>BIFC .....</i>	<i>53</i>
Subcellular fractionation procedure.....	54
Peptide characterization.....	55
<i>Circular dichroism.....</i>	<i>55</i>
<i>Sulforhodamine B leakage assay.....</i>	<i>55</i>
<b>RESULTS .....</b>	<b>57</b>
Chapter 1. Early stages of membrane protein folding .....	59
<i>c.1.1. The glycosylation mapping assay .....</i>	<i>60</i>
<i>c.1.2. TMD but not soluble helices have a compact conformation.....</i>	<i>61</i>

<i>c.1.3. Molecular dynamics simulation helix folding inside the mammalian ribosome</i> .....	65
<i>c.1.4. Determinants of TMD helix folding inside the ribosome</i> .....	66
Chapter 2. Interfacial scale determination.....	71
<i>c.2.1. Experimental design</i> .....	71
<i>c.2.2. pHLIP3D replacement points</i> .....	75
<i>c.2.3. Improvement of the methodology by inserting a third glycosylation acceptor site</i> .....	76
<i>c.2.4. Calculating interfacial propensity for individual amino acids</i> .....	79
<i>c.2.5. Characterization of Bax-derived amphipathic peptide</i> .....	87
Chapter 3. The role of hydrophobic matching on TMD helix packing .....	91
<i>c.3.1. Packing of TMD with different length in biological membranes</i> .....	92
<i>c.3.2. Influence of transmembrane hydrophobic length mismatch on heterotypic helix-helix packing</i> .....	98
<i>c.3.3. Influence of hydrophobic length on the subcellular localization of the chimeras</i> .....	102
<i>c.3.4. Molecular dynamics simulation of homo- and hetero-oligomerization in different size membranes</i> .....	104
Chapter 4. Topological study of MPs.....	109
<i>c.4.1. Membrane topology of the translocon-associated protein gamma subunit (TRAP<math>\gamma</math>)</i> .....	109
<i>c.4.2. Membrane topology of bacteriorhodopsin</i> .....	112
<b>DISCUSSION</b> .....	<b>121</b>
<b>CONCLUSIONS</b> .....	<b>133</b>
<b>RESUMEN</b> .....	<b>137</b>
Introducción .....	139
Objetivos .....	143
Metodología .....	145
Conclusiones .....	151
<b>REFERENCES</b> .....	<b>153</b>
<b>ANNEX I</b> .....	<b>171</b>
<b>ANNEX II</b> .....	<b>175</b>
<b>ANNEX III</b> .....	<b>181</b>
<b>ANNEX IV</b> .....	<b>187</b>
<b>ANNEX V</b> .....	<b>191</b>
<b>ANNEX VI</b> .....	<b>195</b>
<b>PUBLICATIONS DURING THE THESIS</b> .....	<b>199</b>



## List of Figures

Figure 1   Cartoon representation of the secondary structure types of membrane proteins. ....	11
Figure 2   General biological membrane structure.....	12
Figure 3   Structure of a phospholipid bilayer. ....	14
Figure 4   Membrane organization.....	16
Figure 5   Ribosome exit tunnel.....	18
Figure 6   Cartoon depicting the SRP targeting cycle in eukaryotes.....	20
Figure 7   Post-translational model in eukaryotes.....	21
Figure 8   Overall structure of the ribosome-bound mammalian translocon.....	22
Figure 9   Translocon structure.....	23
Figure 10   Sec61 channel pore.....	24
Figure 11   Conformational transitions of Sec61 during co-translational protein translocation and membrane insertion.....	26
Figure 12   Structure of the signal sequence-engaged Sec61 complex.....	27
Figure 13   Sec61 channel pore, subunits $\alpha$ , $\beta$ , and $\gamma$ .....	28
Figure 14   Translocon components and scheme for co-translational N-glycosylation..	29
Figure 15   Structure and subunit composition of the mammalian TRAP complex.....	31
Figure 16   ‘Biological’ amino acid hydrophobicity scale.....	33
Figure 17   TMD amino acid preference.....	34
Figure 18   Amino acid helix conformation.....	35
Figure 19   Some of the possible adaptations to hydrophobic mismatch.....	37
Figure 20   Types of Topogenic determinants of single-spanning membrane proteins..	38
Figure 21   Routes of possible integration of the first two TMDs of membrane proteins.....	40
Figure 22   Amino acid type and position distribution in TMD helices.....	41
Figure 23   Structural representation of membrane proteins.....	42
Figure 24   Cartoon representation of the approach.....	61
Figure 25   Atomic structure of the soluble proteins studied.....	62
Figure 26   Glycosylation profile of assayed sequences.....	63
Figure 27   Insertion of hydrophobic regions of VSV-G, L9, and NAGK into membranes using Lep as model protein.....	63
Figure 28   Glycosylation profile of TMD helices from different origins.....	64
Figure 29   Solvent accessible surface area (SASA) for TMD and soluble sequences in the ribosome exit tunnel.....	66
Figure 30   Solvent accessible surface area (SASA) for folded versus extended states.....	66
Figure 31   Hydrophobicity and helicity affect TMD folding.....	67
Figure 32   Folding depends on hydrophobic length and correlates with insertion.....	68
Figure 33   Needed residues to compact within the ribosome exit tunnel.....	69
Figure 34   Schematic representation of the Lep variants.....	72
Figure 35   Determination of the N- and C-termini minimal glycosylation distance for interfacial sequences.....	73
Figure 36   Interfacial Lep assay for pHLIP3D and melittin sequences.....	74
Figure 37   pHLIP3D sequence screening for systematic substitutions.....	76

Figure 38   Interfacial LepG3 assay.....	77
Figure 39   Amino acid substitutions in DL pair and DLP triplet. ....	79
Figure 40   Schematic representation of the equilibrium among the three possible states. .....	80
Figure 41   Apparent free energy amino acid scale from individual equilibrium constants. ....	82
Figure 42   Amino acid apparent free energy values for each state.....	83
Figure 43   Correlation between biological and biophysical scales. ....	84
Figure 44   Global view for residue's apparent free energy. ....	85
Figure 45   Individual amino acid propensity across the three different states. ....	86
Figure 46   Model of active dimeric bax at the membrane.....	88
Figure 47   BaxE5 selectively disrupts lipid integrity. ....	89
Figure 48   BaxE5 disrupts membranes in a largely unstructured conformation. ....	90
Figure 49   Homo-dimerization in <i>E. coli</i> membranes. ....	94
Figure 50   Schematic representation of BIFC Assay. ....	95
Figure 51   Homo-dimerization in eukaryotic membranes.....	97
Figure 52   Hetero-dimerization in eukaryotic membranes. ....	98
Figure 53   Differences in hetero-dimerization in eukaryotic membranes. ....	99
Figure 54   Heat map representation of BIFC hetero-oligomerization assay. ....	100
Figure 55   Confocal microscopy analysis for membrane dimer formation. ....	101
Figure 56   Confocal microscopy values vs fluorimeter values. ....	101
Figure 57   Homo-dimerization at the endoplasmic reticulum. ....	102
Figure 58   Homo-dimerization at the plasma membrane. ....	103
Figure 59   Sub-cellular localization of full-length GpA.....	104
Figure 60   Stability of the dimeric structures from molecular dynamics simulations. ....	106
Figure 61   Summary of the results from molecular dynamics simulations of homo- dimers. ....	107
Figure 62   Tilt and crossing angle distribution. ....	108
Figure 63   Summary of the results from molecular dynamics simulations of the 17L/29L hetero-dimer. ....	108
Figure 64   Insertion of predicted TMDs into microsomal membranes. ....	110
Figure 65   <i>In vitro</i> analysis of truncated TRAP $\gamma$ constructs.....	112
Figure 66   Predicted Bacteriorhodopsin TMD by $\Delta G$ Predictor server fitted in bR structure. ....	114
Figure 67   <i>In vitro</i> expression and representative SDS-PAGE analysis of short bR truncates.....	115
Figure 68   Glycosylation mapping of ABC truncated version of bacteriorhodopsin. ....	116
Figure 69   Glycosylation mapping of A-D truncates of bacteriorhodopsin. ....	117
Figure 70   Glycosylation mapping of A-E, A-F, and full-length bR.....	118
Figure 71   Glycosylation mapping of bR truncates harboring an extended loop connecting helices C and D. ....	120
Figure 72   Nascent polypeptide conformations within the ribosome exit tunnel. ....	124
Figure 73   Molecular dynamic simulations of homo and hetero-dimers.....	130

## Abbreviations

<b>AHs:</b> Amphipathic helices	<b>MM39:</b> malE- <i>E. coli</i>
<b>BBE:</b> berberine bridge enzyme	<b>MP:</b> membrane protein
<b>BIFC:</b> Bimolecular fluorescence complementation	<b>N<sub>cyt</sub>:</b> N-terminal cytosolic
<b>bR:</b> bacteriorhodopsin	<b>N<sub>lum</sub>:</b> N-terminal luminal
<b>C<sub>cyt</sub>:</b> C-terminal cytosolic	<b>OST:</b> Oligosaccharyltransferase
<b>CD:</b> circular dichroism	<b>PEI:</b> Polyethyleneimine
<b>CF:</b> cytosolic fraction	<b>PK:</b> Proteinase K
<b>C<sub>lum</sub>:</b> C-terminal luminal	<b>PM:</b> plasma membrane
<b>DEPC:</b> 1,2-dierucoyl-sn-glycero-3-phosphocholine	<b>POPC:</b> 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
<b>DLPC:</b> 1,2-dilauroyl-sn-glycero-3-phosphocholine	<b>PPI:</b> Protein-protein interactions
<b>DOPC:</b> 1,2-dioleoyl-sn-glycero-3-phosphocholine	<b>PTC:</b> Peptidyl transferase center
<b>ER:</b> Endoplasmic reticulum	<b>RM:</b> rough microsomes
<b>EYFP:</b> Enhanced yellow fluorescent protein	<b>RNC:</b> ribosome-nascent chain complex
<b>GFP:</b> Green Fluorescent Protein	<b>Sec:</b> SRP-dependent secretase pathway.
<b>GpA:</b> Glycophorin A	<b>SF:</b> subcellular fractionation
<b>GPCR:</b> G-protein coupled receptors	<b>SP:</b> Signal peptidase
<b>LB:</b> lysogeny broth	<b>SR:</b> SRP receptor
<b>LUV:</b> large unilamellar vesicle	<b>SRP:</b> Signal recognition particle
<b>MBP:</b> maltose-binding protein	<b>SS:</b> Signal sequence
<b>MD:</b> Molecular dynamics	<b>TMD:</b> transmembrane domain
<b>MF:</b> membrane fraction	<b>TRAP:</b> Translocon-associated protein
<b>MGD:</b> Minimum glycosylation distance	<b>VC:</b> BIFC-VC155
	<b>VFP:</b> Venus Fluorescent Protein
	<b>VN:</b> BIFC-VN172

## 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
Glutaminc acid	-	Glu	-	E	-	Tryptophan	-	Trp	-	W
Glycine	-	Gly	-	G	-	Tyrosine	-	Tyr	-	Y
Histidine	-	His	-	H	-	Valine	-	Val	-	V
Isoleucine	-	Iso	-	I	-		-		-	

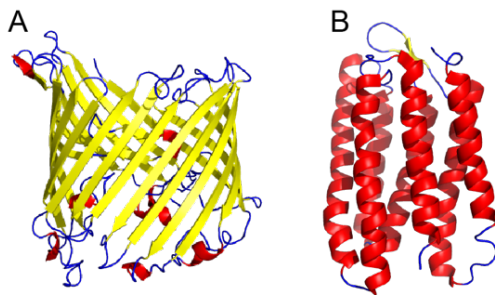


# **INTRODUCTION**



The cell, in biology, is the basic unit of life. It contains the necessary molecules to carry out the homeostasis maintenance. Proteins are among the molecules necessary for the cell. The vast diversity of functions to be carried out by proteins require an accompanying structural and localization diversity. Among all cellular proteins, there is a subset called membrane proteins (MPs). MPs are a type of proteins associated with, or attached to, biological membranes.

Genes encoding for MPs represent about 30% of all human open reading frames (Lander et al. 2001; Venter et al. 2001; Uhlén et al. 2015). They are characterized by the presence of the so-called transmembrane domains (TMD). These TMD's are amino acid sequences adapted to fold, insert, and function in a hydrophobic environment, as provided by the biological membranes. Based on their secondary structure, MPs can be subdivided in two groups.  $\beta$ -barrels, present in the bacteria, mitochondria, and chloroplast outer membranes and  $\alpha$ -helical bundles, found in all biological membranes (Figure 1). In fact, due to their wide membrane distribution, the  $\alpha$ -helical subset of MPs represents approximately 27% of coding proteins in the human genome (Almén et al. 2009). Despite the abundance of MPs, due to the complexity imposed by the environment in which they are found, our understanding of their biosynthesis and folding is far from the knowledge we have of soluble proteins. It is for these reasons that the present thesis will focus on the study of  $\alpha$ -helical MPs biogenesis and folding. Thus, we will start by describing the structure and organization of the biological membrane. Next, we will introduce  $\alpha$ -helical MP synthesis, folding, insertion, and disposition into biological membranes.



**Figure 1 | Cartoon representation of the secondary structure types of membrane proteins.**

(A) Example of  $\beta$ -barrel MP secondary structure, Sucrose-specific porin, PDB code 1A0S. (B) Example of  $\alpha$ -helical MP secondary structure, Bacteriorhodopsin, PDB code 1FFB.  $\alpha$ -helices are red colored whereas  $\beta$ -sheets are yellow colored.

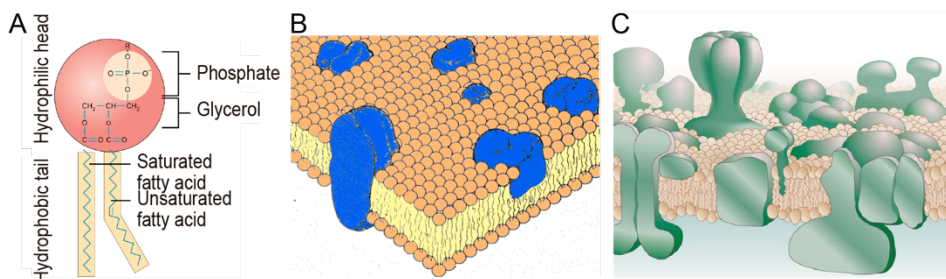
## I.1. Biological membranes

The biological membranes are the barriers that define and delimit living cells and the organelles within. They are fundamental to maintain cellular structural integrity and organization. There are two main components in biological membranes, lipids (predominantly phospholipids (Figure 2A)) and proteins. Structurally, the lipids, in a biological membrane, are organized in two opposing layers forming a continuous

## Introduction

barrier, also known as bilayer. Concerning proteins, they can be classified based on their physical arrangement with respect to the membrane as integral (that is inserted into the membrane) and peripheral (associated with the membrane surface) MPs. Without MPs, the membrane would be a hermetic barrier, unable to mediate cell transport, communication, signaling, or energy production.

One of the early descriptions of a biological membrane that considered the presence of lipids and proteins was the fluid mosaic model (Singer & Nicolson 1972). In this model, Singer and Nicolson presented a membrane in which MPs can be found in the lipid bilayer as boats floating in an ocean of lipids (Figure 2B). More recent studies have refined the model, adding new particularities, primarily high protein concentration (up to  $\sim 30,000$  proteins/ $\mu\text{m}^2$  in some membranes) with a non-random distribution mediated by several lipid-protein and protein-protein interactions (PPI). These particularities lead to propose a model for the biological membranes where segregated regions with particular structures and functions could be envisioned (Quinn et al. 1984; Engelman 2005; Dupuy & Engelman 2008) (Figure 2C).



**Figure 2 | General biological membrane structure.**

(A) Cartoon representation of a phospholipid. Structurally, a phospholipid can be divided in two parts, the head and the tail. The head, hydrophilic in nature, is formed by glycerol, a phosphate group, and a substituent (headgroup) attached to the glycerol. The hydrophobic tail contains two fatty acids, which can include one or more double bond in their aliphatic chains. (B) The Singer-Nicolson ‘fluid mosaic model’ where the authors presented a membrane in which MPs can be found in the lipid bilayer as boats floating in an ocean of lipids (adapted from (Singer & Nicolson 1972)). (C) An updated view of the membrane structure where MPs segregate in regions at higher concentration than proposed in previous models (taken from (Engelman 2005)).

### i.1.1. Structure of biological membranes

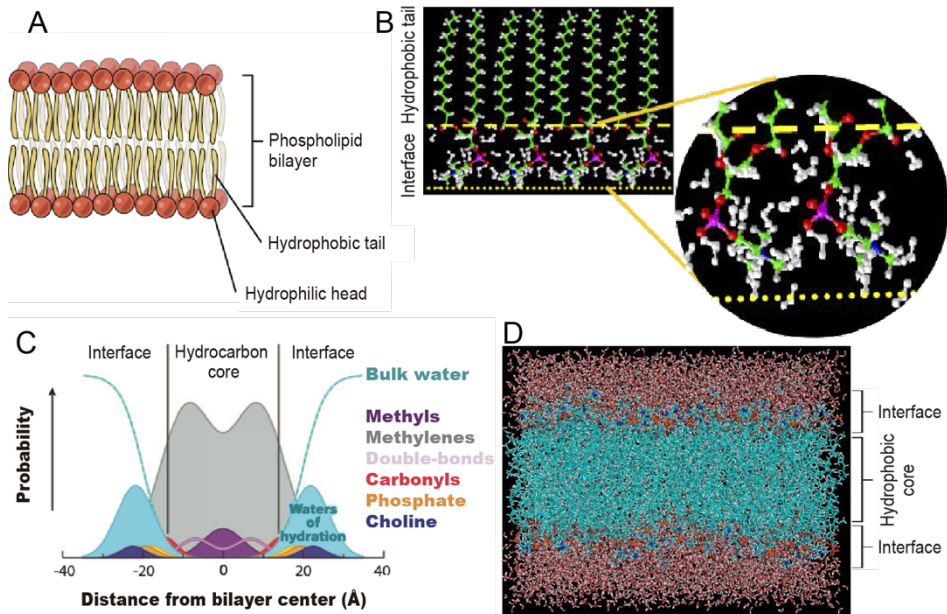
Structurally, biological membranes are primarily determined by their lipid composition. Although lipid molecules display considerable structural diversity, they share one important property, the amphipathicity. Amphipathic molecules are characterized by possessing both a hydrophilic and hydrophobic nature. The



combination of both properties in the same lipid molecule drives the formation of organized structures when placed in an aqueous environment, which is the basis for the lipid organization in biological membranes. Biological membranes are organized in two leaflets of lipids whose hydrophilic regions (e.g., the phosphate-containing headgroups) are facing towards the aqueous environment while, the most hydrophobic regions (the hydrocarbon tails) face the interior of the membrane where they are protected from the aqueous environment (Figure 3A). This lipid organization presents a hydrophobic core, where the hydrocarbon tails are facing the tails of the other monolayer, and a membrane frontier where the hydrophilic headgroups of the lipids form a transition zone (also known as membrane interface) towards the bulk water (Figure 3B). The interfacial compartment created by the headgroups on each monolayer provides a heterogeneous chemical region (Figure 3B). This membrane interface allows a gradual change zone between the aqueous environment and the hydrophobic core, in which the water is organized in different levels on the surface (Disalvo et al. 2008).

The dimensions of a lipid bilayer were determined in 1992 by Wiener and White using dioleoylphosphatidylcholine (DOPC) as a model lipid. On average, the hydrophobic core spreads 30Å of thickness, while the interface region spread 15 Å on each side of the bilayer (Figure 3C) (Wiener & White 1992). Therefore, the hydrophobic core and both interface regions together are comparable in size (Figure 3D). However, both lipid headgroups and tails can differ in carbon length and number of saturations in actual membranes, which can directly affect the hydrophobic core and interface thickness, and thus modulate the properties of biological membranes (van Meer et al. 2008).

## Introduction



**Figure 3 | Structure of a phospholipid bilayer.**

(A) Cartoon representation of the lipid disposition in a phospholipid bilayer. The lipids are placed in two leaflets whose phosphate-containing headgroups (hydrophilic head) are facing towards the aqueous environment protecting the hydrophobic tail from the bulk water. (B) Computational modelation of the membrane interface, where a gradual change between the aqueous environment and the hydrophobic core takes place. The region between the two planes (yellow dotted lines) confines the hydrated groups and water. Taken from (Disalvo et al. 2008). (C) Gaussian probability distribution of the lipid component groups of a DOPC model bilayer. The areas under the curves correspond to the number of constituent groups per lipid represented by the distributions. Taken from (White & G. von Heijne 2008; Wiener & White 1992). (D) Simulated box of POPC all-atom model membrane. Both interfaces and the hydrophobic core are comparable in size. Taken from (Stachura & Kneller 2015).

### i.1.2. Membrane organization

Biological membranes are far from being static and homogeneous structures. As previously mentioned, the large chemical diversity found in lipid headgroups and acyl chains introduce variability in the membrane structure, which in turn modulates the membrane properties. But not only the lipid diversity affects the membrane properties, but the presence of MPs also affects membrane function and structure.

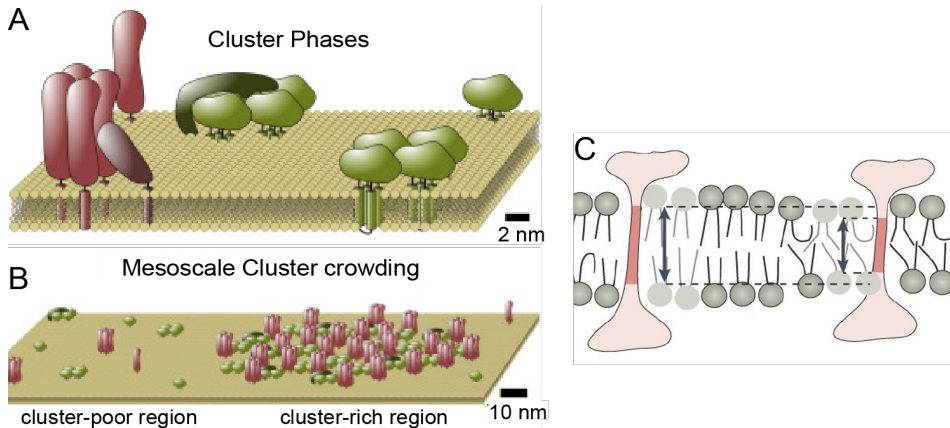
Studies in synaptic vesicles (Takamori et al. 2006) and red blood cell membranes (Dupuy & Engelman 2008) determined that 18-25% of the hydrophobic volume of

the membrane is occupied by TMDs. The MPs to which the TMDs belong are not randomly distributed but they are clustered in domains that compartmentalize cellular processes as the lipid raft hypothesis postulated (Figure 4A) (Simons & Ikonen 1997; Pike 2006).

In recent years, a new membrane organization model has been proposed (Destainville et al. 2016). In this model, MPs are also grouped in clusters that do not randomly scatter across the membrane with the main difference that these clusters are grouped into a larger organization that must be viewed from a mesoscale perspective (Figure 4B). Several studies evidenced that most of the proteins segregate into these different domains or clusters via specific mechanisms such as cholesterol-enriched domains or rafts (Chamberlain et al. 2001; Lang 2007),  $\text{Ca}^{2+}$  concentration (Zilly et al. 2011), or hydrophobic mismatch (Milovanovic et al. 2015). Among all membrane properties that affect its organization the present thesis is going to focus on the hydrophobic thickness of the membrane, particularly in the hydrophobic mismatch between the membrane and the TMDs.

Hydrophobic thickness is the length of the lipid bilayer's hydrophobic core. The hydrophobic thickness is influenced not only by the number of carbons or the saturation of their tails, i.e. lipids with long and saturated fatty acids as sphingolipids (Sezgin et al. 2017) but also by the phospholipid/esterol ratio (Nezil & Bloom 1992; Weiss et al. 2003; Mitra et al. 2004). Hydrophobic mismatch occurs when the length of the hydrophobic region of a TMD does not match with the membrane's hydrophobic thickness (Figure 4C). This mismatch would expose hydrophobic amino acid to the aqueous environment or polar residues to the hydrophobic core of the membrane and thus is usually avoided by matching the hydrophobic thickness of the TMDs with the hydrophobic core length of the lipids around them. Several studies of TMDs datasets have shown that the distribution of MPs along different organelles depends on the membrane thickness, being MPs with longer TMDs preferably found at the plasma membrane (PM), whereas shorter TMDs remain attached in the endoplasmic reticulum (ER) and Golgi membranes, thinner than the PM (Sharpe et al. 2010; Singh & Mittal 2016). Additionally, some studies support the idea that lipids and proteins may not be naturally matched, but rather TMD length shapes the bilayer thickness by modulating lipid acyl chain conformation and packing by means of the hydrophobic effect and specific lipid-protein interactions (Nezil & Bloom 1992; Weiss et al. 2003; Mitra et al. 2004). These last adaptation mechanisms to reduce hydrophobic mismatch are revised in i.3.3.

## Introduction



**Figure 4 | Membrane organization.**

(A) Cluster phases model. The model suggests the concentration of only one or a small group of proteins into a densely packed protein cluster. The clusters are separated from each other through a protein-free zone. (B) Mesoscale organization of the cluster phase. Clusters from different proteins clump together into a large mesoscale assembly. Taken from (Destainville et al. 2016). (C) Mismatch between the length of TMD and membrane thickness causes lateral displacement of the protein to adjust the TMD length with the bilayer thickness. Taken from (Harayama & Riezman 2018).

## I.2. Membrane protein synthesis and targeting

Protein synthesis occurs in the cytosol where the mRNA is translated by ribosomes. The process advances by adding a codon-specific amino acid to the growing polypeptide chain. As new amino acids are added, the nascent polypeptide chain navigates through the ribosome, emerging at the exit site. All proteins must acquire their native conformation during or after protein translation to complete their functions. Soluble proteins fold by favorable thermodynamic paths until the final lowest energy state (native state) is achieved, a process that might require chaperone assistance to reach their final folded state in the cytosol.

The situation for secretory and MPs is significantly different. They have to cross or insert into the membrane, which means that they have to overcome the thermodynamic barrier that the biological membrane imposes. Additionally, the complete folding of the protein is not allowed prior to membrane insertion or translocation, due to space restrictions in the machinery involved in these processes. All in all, it indicates that chaperones also play a key role during secretory and MPs translocation/insertion and folding. In the next sections, we will revise all the steps and cellular machinery necessary to achieve the native structure of a MP, from the initial folding to the final TMD insertion process.

### i.2.1. The ribosome

In all cells, ribosomes translate the information encoded in mRNAs into proteins. Ribosomes are large ribonucleoproteins composed of two-thirds of rRNA and one-third of proteins. They present two subunits of different sizes. The small subunit is in charge of the fidelity of the process, mediating the interaction between the tRNA anticodons and the mRNA codons (Garrett et al. 2000). The large subunit, on the other hand, catalyzes the formation of the peptide bond and mediates the binding of multiple factors that assist during the translation process.

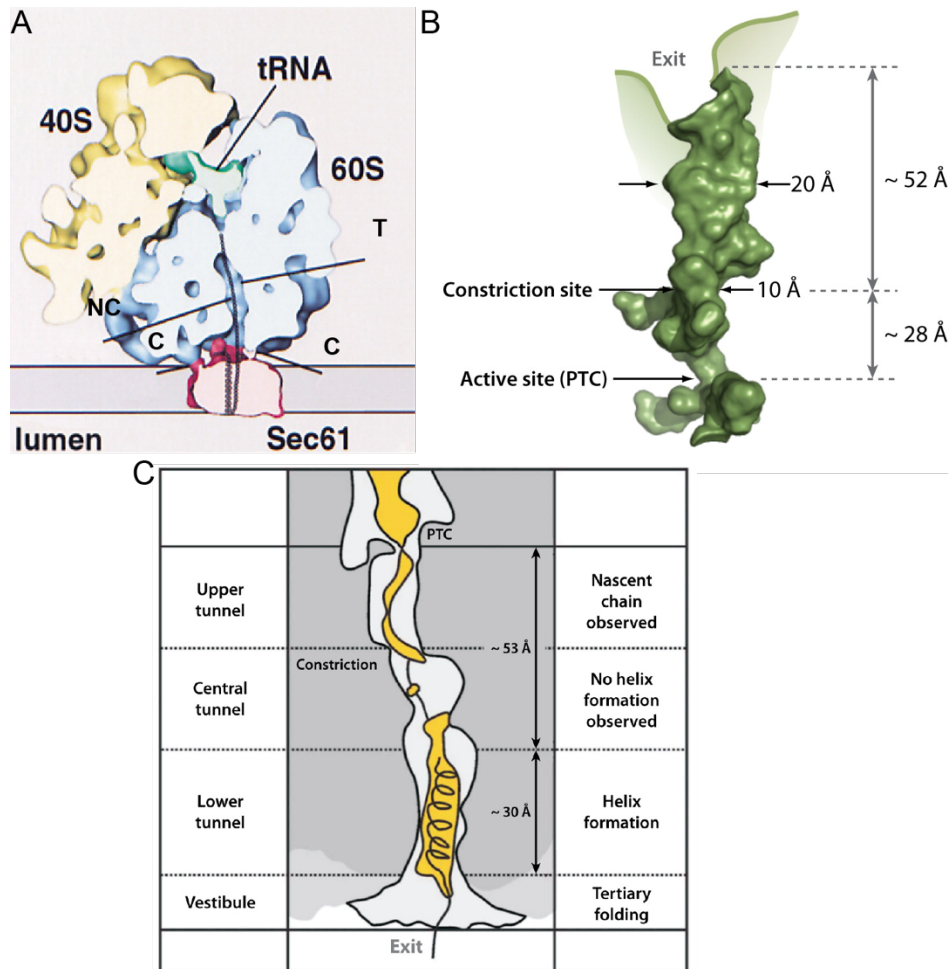
#### *i.2.1.1. The ribosome exit tunnel*

The peptide bond formation is catalyzed in the peptidyl transferase center (PTC) placed at the large ribosomal subunit. After peptide bond formation the polypeptide chain is released from the PTC towards the ribosome exit site by moving through a connecting channel known as ribosome exit tunnel. The ribosome exit tunnel accommodates the nascent polypeptide chain as it navigates from the PTC site to the ribosomal exit situated at the end of the subunit (Frank et al. 1995) (Figure 5A). CryoEM approaches allowed the structural characterization of the ribosome exit tunnel, which is mainly formed by rRNA and the ribosomal proteins L4, L17, and L39 in eukaryotic cells (L4, L22, and L24 in bacteria) (Frank et al. 1995; Ban et al. 2000; Bhushan et al. 2010). The length of this tunnel is around 80-100 Å (depending on where the exit-side is defined), which provides enough space to accommodate 30-40 amino acids of the nascent chain traveling through it (Malkin & Rich 1967; Beckmann et al. 2001). The tunnel width ranges between 10 and 20 Å. The narrowest part of the tunnel is placed ~28Å away from the PTC. This constriction is caused by a  $\beta$ -hairpin loop of the ribosomal proteins L17 (L22 in bacteria) and L4 (Voss et al. 2006) (Figure 5B).

The constrained environment of the ribosome exit tunnel provides a passive protective shield for the nascent chain (Cassaignau et al. 2020). In recent years, newly acquired data indicates that the tunnel has an active role in the folding of some nascent chains (Mingarro et al. 2000). First, it has been discovered that the exit tunnel provides enough space to allow the folding of secondary structures as  $\alpha$ -helices. Precisely, FRET assays (Woolhead et al. 2004) revealed that nascent chains far inside the tunnel could adopt an  $\alpha$ -helix-like structure. Later, Cryo-EM experiments and single-particle reconstitution (Bhushan et al. 2010) enabled direct visualization of a structure compatible with an  $\alpha$ -helix inside the ribosome exit tunnel. Furthermore, these assays suggested interactions between the tunnel wall and the compacted nascent chain (Figure 5C). Recently, it has been shown that not only secondary

## Introduction

structure can be acquired inside the ribosome tunnel. Small domains with tertiary structures such as zinc-finger domain have been observed folded close to the exit site, where the ribosome tunnel widens substantially (Hedman et al. 2015) (Figure 5C).



**Figure 5 | Ribosome exit tunnel.**

(A) Representation of a ribosome-nascent chain-Sec61 (RNC-Sec61) complex cut perpendicularly to the plane of the membrane along the tunnel in the large ribosomal subunit (blue colored). In this panel, C indicates the contact site between the ribosome and the Sec61(magenta) channel. The position of the ribosome exit tunnel (T) and the nascent chain (NC) are shown. Taken from (Beckmann et al. 2001). (B) Structure of the ribosomal exit tunnel. Peptidyl transferase center (PTC). Constriction site and tunnel widening are marked with arrows. Adapted from (Voss et al. 2006). (C) Schematic cross-section of ribosomal large subunit after translating helix1 (5x EAAAK repeats) that represents the helix and tertiary structure formation within the lower tunnel and the vestibule of the exit tunnel. Taken from (Bhushan et al. 2010).

### i.2.2. Targeting to the Sec translocon in eukaryotic cells

Once the nascent MP chain emerges from the ribosome, protein targeting to the ER membrane can occur co-translationally or post-translationally. The Signal Recognition Particle (SRP) plays a central role in the co-translational pathway, where ribosome-nascent chain (RNC) complexes are captured, translation halted, and addressed to the ER membrane at the beginning of the translation process. Other components of the co-translational route include the SRP receptor (SR) and the Sec61 translocon, which forms a pore through the ER membrane in eukaryotic cells. Assisted by these components the elongating polypeptide chain is released directly from the ribosome into the translocon channel (Rapoport 2007).

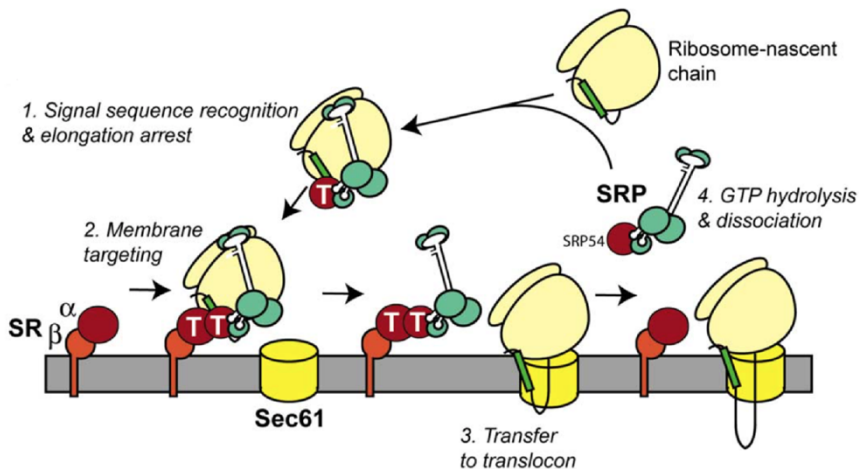
In the post-translational pathway, proteins are targeted and inserted (or translocated) once translation by cytosolic ribosomes has ended. A dedicated complex including Sec62-Sec63 proteins cooperates with the Sec61 translocon to insert or translocate these fully synthesized proteins. Proteins involved in this route require a chaperone-assisted maintenance of the partially unfolded state to subsequently be able to cross the lipid bilayer. Additionally, the post-translational targeting of a group of MPs that present a C-terminal hydrophobic region, also known as tail-anchored (TA) proteins, is carried out by the TRC40-Get pathway (detailed explanation in i.2.2.1).

Whether a protein uses the co-translational or post-translational pathway depends on the presence of a signal sequence (SS) and the location of the hydrophobic regions in the case of MPs. The SS consists of a short stretch of hydrophobic residues flanked by positively charged amino acid residues and polar but uncharged residues at the N- and C-terminal end respectively (Cross et al. 2009). Once the protein has reached its destination the SS is cleaved from the polypeptide to render the mature protein.

During co-translational targeting, the ribonucleoprotein SRP complex binds to the SS as it emerges from the ribosome (Pool 2005). In the absence of SS the first TMD can act as a non-cleavable SS. Once the SRP has bound the RNC complex, protein synthesis is transiently arrested and the RNC-SRP complex is docked to the ER membrane via the SR. The SR is a hetero-dimer formed by two GTPases, SR $\alpha$  and the membrane-anchored SR $\beta$ . The interaction between SRP and SR requires GTP binding to both complexes. Next, the RNC complex is transferred to the translocon, a process that requires the hydrolysis of the SR-bound GTP which in turn triggers SRP-SR dissociation (Song et al. 2000) (Figure 6).

## Introduction

Binding of the RNC complex to the translocon aligns the ribosome exit tunnel with the translocon pore (Beckmann et al. 1997). After SRP dissociation from the RNC complex translation restarts and the nascent polypeptide is released through the translocon channel, allowing the translocation of secreted proteins and the insertion of MPs into the ER membrane.



**Figure 6 | Cartoon depicting the SRP targeting cycle in eukaryotes.**

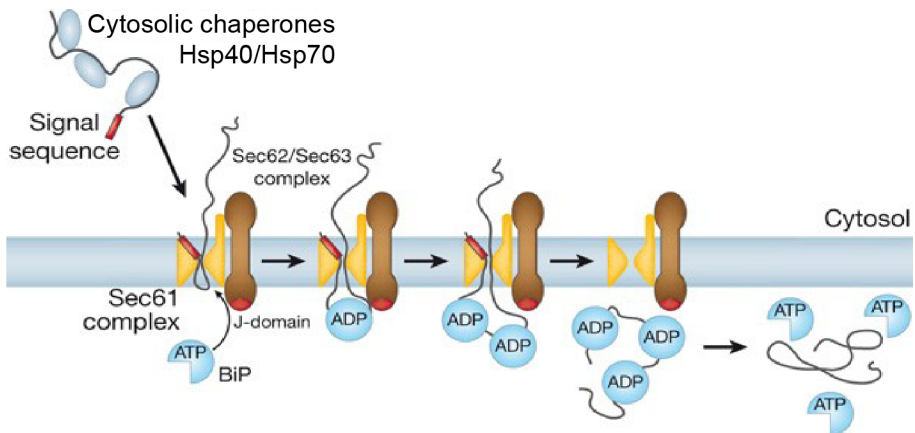
As the SS emerges from the ribosome it is bound by the SRP, a process that arrests translation (1). Next, the RNC complex binds the SR (2). Docking of SRP with SR requires GTP (T) binding by SRP54 and SR $\alpha$ , being both GTPases. Once the signal sequence is transferred to the Sec61 translocon (3), both the SRP and SR hydrolyze GTP and dissociate, allowing the ribosome-bound translocon complex to restart translation (4). Taken from (Pool 2005).

The role of the Sec translocon is not restricted to co-translational translocation as originally thought. The translocon is also involved in the post-translational insertion/translocation of proteins (Wu et al. 2019). Therefore, when a protein is inserted/translocated once it is fully translated, it is necessary to prevent its premature folding since the folded protein cannot fit through the translocon channel.

In eukaryotes, a dedicated complex (Sec62-Sec63 complex) cooperates with the Sec61 translocon to mediate the post-translational insertion or translocation (Wu et al. 2019). In this pathway, the targeted proteins are MPs or secreted proteins preceded by poorly hydrophobic SS that escape the SRP recognition. In this case, the cytosolic Hsp40/Hsp70 type chaperones are in charge of maintaining these polypeptides in a translocation-competent unfolded state and transfer them to the Sec62/Sec63 complex. As soon as the protein is captured by the Sec62/Sec63 complex it is transferred to the Sec61 translocon. Once the polypeptide is in the translocon channel chaperones are released (Figure 7). In addition to the Sec62/Sec63 complex, to carry out the process of insertion/translocation of the polypeptide the luminal ATP-



dependent BiP chaperone is required. BiP binds the polypeptide in the lumen as a result of ATP hydrolysis, preventing the polypeptide from returning to the cytosol. Once the polypeptide is bound the BiP chaperone pull from it towards the ER. BiP acts as a motion force in the post-translational insertion/translocation process since no pushing forces are present by the lack of translating ribosomes. Finally, the exchange of ADP for ATP in the lumen opens the BiP peptide-binding pocket, releasing the whole translocated polypeptide or the luminal portion of a MP.



**Figure 7 | Post-translational model in eukaryotes.**

Cytosolic chaperones HSp40/Hsp70 bind to the polypeptide and associate to Sec62/Sec63 for further polypeptide chain release to the Sec61 complex for translocation. The luminal BiP chaperone mediates the polypeptide retention in the ER lumen by peptide binding as a result of ATP hydrolysis by the J-domain of Sec63. BiP acts as a pulling force that slides the polypeptide chain within the Sec61 complex toward the ER lumen. ADP to ATP reconstitution in the ER lumen enables the release of BiP from the polypeptide. Adapted from (Rapoport 2007).

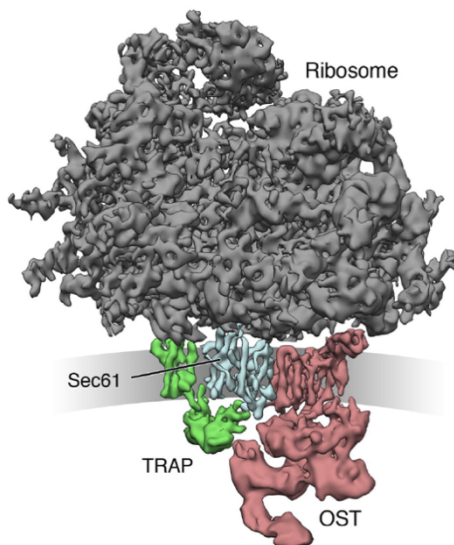
### *i.2.2.1. Tail-anchored MPs' targeting*

Post-translational targeting of C-terminal TMD proteins, also known as TA proteins, does not follow the SRP or the Sec62/Sec63 pathways but the TRC40-Get pathway (Transmembrane Recognition Complex subunit of 40 kDa in mammals or Guided Entry of Tail-anchored proteins in yeast) (Borgese & Fasana 2011). This pathway has been proposed to be independent of the Sec61 complex. The critical point of a TA protein targeting is its selective and efficient recognition by TCR40 (or Get) (Stefanovic & Hegde 2007). The interaction between TCR40 and the TMD shields the hydrophobic C-terminal region of the TA proteins as it transits through the cytosol. Next, the TRC40-TA protein complex must be targeted to the ER where TA proteins are inserted in a, supposedly, Sec61 complex independent manner (Shao & Hegde 2011).

## Introduction

### i.2.3. The Translocon

The eukaryotic translocon machinery is responsible for the insertion of the vast majority of proteins into the ER membrane or, alternately, for the secretion of proteins to the ER lumen. As previously described, protein translocation or insertion either co- or post- translational occurs through the translocon, which means that this machinery is implicated in handling nearly all the non-cytosolic proteins. The eukaryotic translocon machinery includes the Sec61 complex, the oligosaccharyltransferase (OST) complex, the signal peptidase (SP), and several accessory proteins (Figure 8). In this section, we are going to describe the structure and function of each of these components.



**Figure 8 | Overall structure of the ribosome-bound mammalian translocon.**

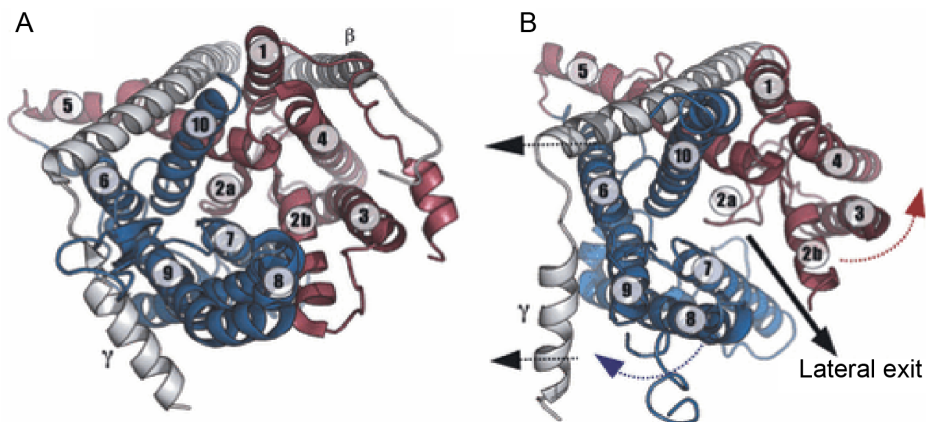
Tomographic densities for the ribosome (gray) and the main translocon constituents, Sec61 complex (light blue), TRAP (translocon-associated protein, green), and OST (red). Taken from (Pfeffer et al. 2016).

#### *i.2.3.1. The Sec61 complex*

The Sec61 (or SecYEG in bacteria and archaea) complex is a hetero-trimeric membrane protein complex with three sub-units (Sec61 $\alpha$ , Sec61 $\beta$ , and Sec61 $\gamma$  in mammals). The  $\alpha$ - and  $\gamma$ - subunits are required for cell survival, showing not only similar structure in both the Sec61 and SecYEG complexes but also significant sequence conservation (Van den Berg et al. 2004; Voorhees et al. 2014). On the contrary, the  $\beta$ -subunit is not essential for cell viability and does not have a significant bacteria and eukarya sequence homology.

## Sec61 $\alpha$

The Sec61 $\alpha$  is the main component of the translocon complex, being the larger (53 kDa) subunit. The  $\alpha$ -subunit is an  $\alpha$ -helical MP formed by 10 TMDs, with both N- and C-termini facing the cytosol (Wilkinson et al. 1996). The Sec61 $\alpha$  structure consists of 2 pseudo-symmetrical N- and C-terminal halves (each of them comprising 5 TMD) forming a channel, a passive conduit through which nascent chains can partition into or translocate across the ER membrane (Figure 9). During co-translational insertion/translocation when a nascent chain enters to the translocon, the Sec61 tunnel is aligned with the ribosomal exit tunnel. This interaction between the ribosome and the translocon provides a protective shield for the transit of the nascent chain from the ribosome to its final destination (the ER membrane for those segments that will be inserted or the ER lumen for translocated domains) (Becker et al. 2009; Voorhees et al. 2014).



**Figure 9 | Translocon structure.**

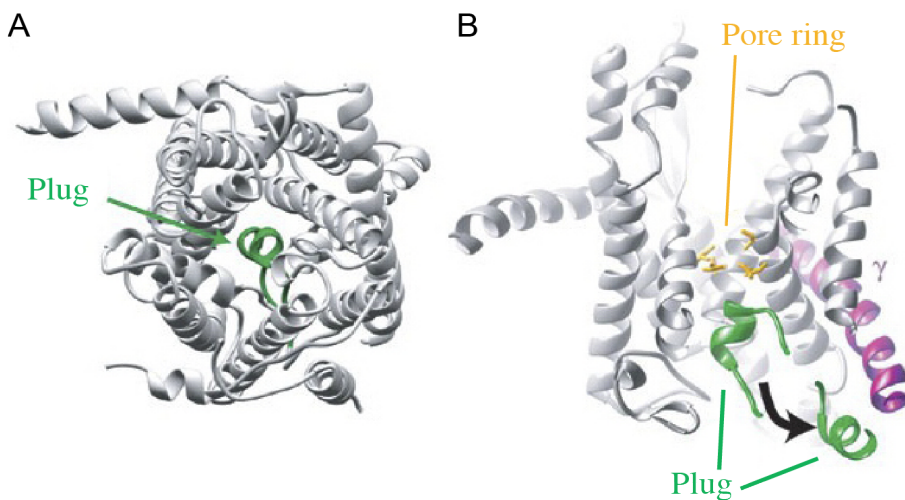
Top view of the translocon structure. (A) Inactive and closed structure of the translocon from *Methanococcus jannaschii*, PDB code 1RHZ (Van den Berg et al. 2004). The translocon channel is blocked by the TMD2a, necessary to maintain the ion permeability barrier. (B) Partially open structure of the translocon from *Pyrococcus furiosus*, PDB code 3MP7 (Egea & Stroud 2010). In both panels, all TMD of Sec61 $\alpha$  are colored (red and blue for each half; see text) except for the  $\beta$ - and  $\gamma$ -subunits, which are shown in gray. All TMDs 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 TMD from the interior of the channel into the ER membrane. Taken from (Martínez-Gil et al. 2011).

A lateral view of Sec61 $\alpha$  reveals a rectangular contour within an hourglass shape. From a top view, the cytoplasmic entry to the Sec61 $\alpha$  channel has a diameter of 20-25 Å, reaching the narrowest point (5-8 Å in diameter) close to the middle of the membrane. This constriction is composed of a ring of bulky hydrophobic residues

## Introduction

followed by a short helix (TMD2a, also known as ‘plug’) that acts as a pore channel blocker in the idle translocon (Figure 10A). After the plug, the channel widens again towards the ER lumen (Van den Berg et al. 2004) (Figure 10B). The Sec61 $\alpha$  subunit also presents a lateral opening, known as lateral gate. The lateral gate is formed by two adjacent TMDs from each Sec61 $\alpha$  halves (TMD2b and TMD7). The displacement of TMD2b and TMD7 is linked to the partition of a TMD into the lipid phase for its final membrane insertion (Van den Berg et al. 2004; Egea & Stroud 2010) (Figure 9B).

When a polypeptide chain reaches the translocon tunnel, two functional scenarios of Sec61 (open and closed states) are presented when it comes to the lateral gate. The open state is characterized by the displacement of the TMD2b and TMD7, allowing sufficiently hydrophobic helices to partition into the ER lipid bilayer. Contrarily, the closed state is characterized by a more compact conformation with the lateral opening closed, thus facilitating the translocation of the polypeptide chain (Van den Berg et al. 2004; Gogala et al. 2014).



**Figure 10 | Sec61 channel pore.**

(A) View from the top of Sec61 channel. The TMD2a (plug, dark green highlighted) acts as a pore channel blocker. (B) View from the side of the Sec61 channel with the front half of the model cutaway. Hourglass shape can be visualized. The arrow indicates the plug (green) movement towards the  $\gamma$ - subunit (magenta) in an open channel state. The hydrophobic pore ring (gold) is shown by the side chains. Taken from (Van den Berg et al. 2004).

Therefore, different three conformations can be distinguished depending on the presence or absence of a polypeptide chain and the nature of the polypeptide chain regarding the presence of TMDs (Gogala et al. 2014).

For idle Sec61 complex, where no polypeptide chain is passing through it, the translocon is found in an inactive conformation. As previously stated, this scenario is defined by a closed lateral gate and a central constriction of the channel, where the TMD2a together with the TMD10 block the tunnel, separating the cytoplasmic side from the external aqueous space (Figure 11A).

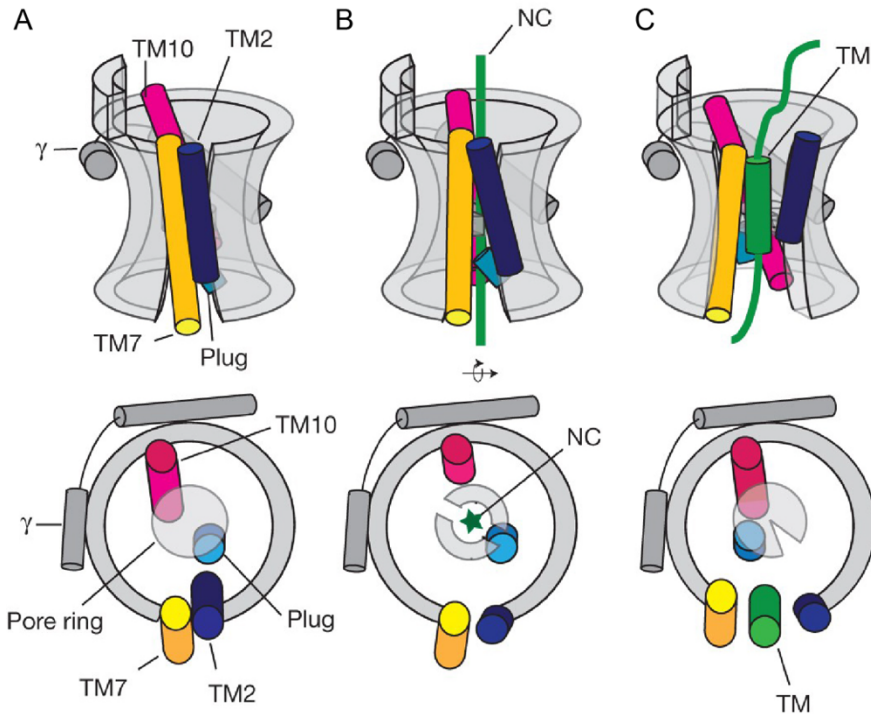
When a translocating polypeptide chain is accommodated in the Sec61 channel, the plug remains stable as in the idle Sec61 complex, but the luminal part of TMD10 shifts outward, away from the plug ( $\sim 6$  Å). This movement would be sufficient to fit an extended translocating peptide segment between the plug and TMD10 (Cannon et al. 2005; Gogala et al. 2014). Concerning the lateral gate, it is partially opened by a few Ångstroms (4 Å) due to a lateral shift of Sec61 $\alpha$  TMD7 (Plessis et al. 2009; Gogala et al. 2014) (Figure 11B).

When the Sec61 is engaged with a more hydrophobic peptide stretch supposed to be inserted into the lipid bilayer as a TMD, the lateral gate opens as a consequence of the displacement of TMD2 and TMD7 ( $\sim 12$  Å). That gap enables the putative TMD to partition into the lipid bilayer (Plath et al. 1998; Van den Berg et al. 2004; Gogala et al. 2014). The TMD10 maintains the same spatial disposition that was viewed in the idle Sec61 complex. The plug, conversely, moves towards the central constriction of the channel closing the void created by the separation of the halves (to allow the lateral gate opening) and maintaining a sealed central channel when the hydrophobic polypeptide arrives (Lycklama a Nijeholt et al. 2011; Gogala et al. 2014) (Figure 11C).

It has also been proposed that the spatial location of the TMD2b-TMD7 in the closed state is consistent with a role for the SS as a trigger for channel opening. In this scenario, the SS would have access to its binding site in the closed channel. The binding modifies the interaction of the plug, keeping the plug in the center of the pore and preventing it from returning to its closed-state position (Voorhees & Hegde 2016) (Figure 12). The final cleavage of the SS by the signal peptidase then will allow TMDs to partition into the lipid phase (Heinrich et al. 2000).

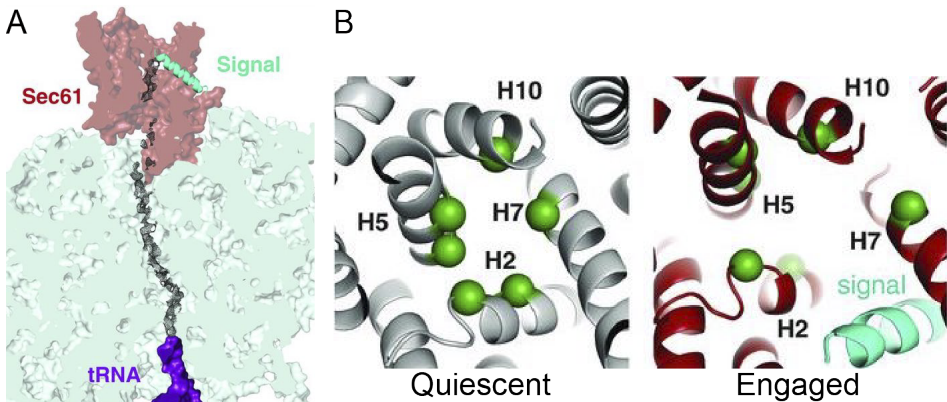
The Sec61 pore diameter has also been reported to undergo a significant increase during MPs insertion. It probably occurs due to the need to accommodate multiple TMD in multi-spanning nascent chains that may leave the translocon in pairs or groups to facilitate membrane integration (Hamman et al. 1997).

## Introduction



**Figure 11 | Conformational transitions of Sec61 during co-translational protein translocation and membrane insertion.**

(A) In the ribosome-bound idle state, both the lateral gate and the central constriction of the Sec61 complex are closed by the TMD2 and TMD7 and TMD10 and the plug, respectively. (B) When the Sec61 complex is engaged with a translocating peptide, the luminal part of TMD10 moves outward creating a central opening for the hydrophilic nascent chain (green star). The plug and TMD7 remain unchanged. The TMD2 rearranges its position slightly, resulting in a partial opening of the lateral gate. (C) Finally, when the Sec61 is engaged with a more hydrophobic peptide stretch, that is supposed to be inserted into the lipid bilayer as a TMD, the lateral gate opens up further. The aperture could accommodate a peptide segment in a helical conformation between TMD2 and TMD7 to allowing access to the lipid phase. Lateral gate opening and transfer of the hydrophobic peptide from the central aqueous channel into the lateral gate are accompanied by a concomitant inwards movement of the plug and TMD10 to maintain the ion permeability barrier of the ER membrane. Taken from (Gogala et al. 2014).



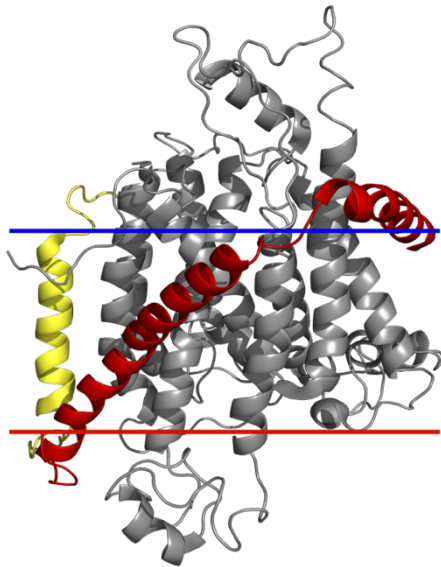
**Figure 12 | Structure of the signal sequence-engaged Sec61 complex**

(A) Density observed from the nascent polypeptide through the ribosomal exit tunnel and parts of the Sec61 channel. The SS (cyan helix) is placed in the lateral gate of the Sec61 (red colored) (B) View of the pore ring residue positions (green spheres) in the quiescent SecY crystal structure (gray, PDB code: 1RH5) and engaged Sec61 complex (red). Signal sequence is colored in cyan. Taken from (Voorhees & Hegde 2016).

Recent cryo-electron tomography and subtomogram averaging experiments revealed that the conformation of Sec61 complex with the lateral exit opened may be the conformation present when the ribosome is bound to the translocon. These results suggest that the ribosome alone (even without a nascent chain) is sufficient to induce a partially open translocon conformation (Pfeffer et al. 2015), which would support a ‘sliding model’. In this model, the nascent chains are being exposed to lipids, due to the laterally ‘pre-opened’ conformation. Thus, depending on the hydrophobicity of the polypeptides, and following the lowest-energy trajectory, they would either move across the translocon into the ER lumen or partition into the lipid bilayer (Cymer et al. 2015). Note that, although the laterally open conformation might be the most predominant in ribosome-translocon complexes, the closed conformation is found in ribosome-free idle state Sec61 complexes, maintain the necessary ion permeability barrier.

## Sec61 $\beta$

The Sec61 $\beta$  subunit is a small (9.9 kDa) membrane protein with a single TMD, located near the TMD1 and TMD4 of Sec61 $\alpha$  (Figure 13). Sec61 $\beta$  is a non-essential component of the translocon. Nonetheless, it has been reported to kinetically facilitate cotranslational translocation (Kalies et al. 1998) and to be implicated in the recognition of unoccupied translocons (Jiang et al. 2008).



**Figure 13 | Sec61 channel pore, subunits  $\alpha$ ,  $\beta$ , and  $\gamma$**

Cryo-EM structure of the Sec61-complex bound to the large ribosome subunit (not shown) translating a membrane-inserting substrate. PDB code 4CG6 (Gogala et al. 2014). Subunits are colored in gray (Sec61 $\alpha$ ), yellow (Sec61 $\beta$ ), and red (Sec61 $\gamma$ ). Red (extracellular) and blue (cytoplasmic) lines indicate approximate membrane location calculated using the PPM server (M. A. Lomize et al. 2012).

### Sec61 $\gamma$

The Sec61 $\gamma$  is the smallest component (7.7 kDa) of the complex. Despite its low molecular weight, it is crucial for the maintenance of the complex structure. Sec61 $\gamma$  is composed of two  $\alpha$ -helices linked by an extended loop (Figure 13). The first helix contains multiple charged residues (positive and negative) that facilitates an amphipathic positioning parallel to the cytosolic side of the membrane, while the second helix spans the bilayer as a TMD (Gogala et al. 2014). Sec61 $\gamma$  interacts with both the N- terminal and C-terminal halves of the Sec61 $\alpha$  and clamps them together (Figure 13) (Van den Berg et al. 2004).

#### *i.2.3.2. The Oligosaccharyltransferase complex*

Protein processing in the secretory pathway is essential for proteins destined to be secreted from the cell or integrated into cellular membranes. More than 70% of all proteins that are processed by ER, Golgi, and vesicles implied in the secretory pathway are glycosylated (Gavel & G. V. Heijne 1990). The complexity of the oligosaccharides which can be found in processed glycoproteins is astonishing. This complexity is due to the variety of chemical linkages that can be formed to a single saccharide building block, adding complexity to the glycoprotein structures.

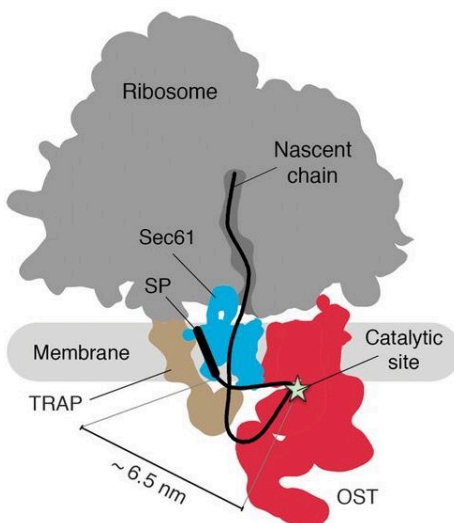
In N-linked glycosylation, a highly conserved post-translational protein modification, the OST (also known as dolichyl-diphosphooligosaccharide-protein glycotransferase) is the enzyme that catalyzes the saccharide transfer. The enzyme



transfers a tetradecasaccharide molecule ( $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ ) from a dolichol pyrophosphate donor to an Asn side chain acceptor in the nascent polypeptide chain. Only the Asn residue located within certain amino acid sequence act as efficient glycosylation acceptor sites. The glycosylation acceptor sites are determined by the amino acid sequence Asn-Xaa-Ser or Asn-Xaa-Thr, where Xaa is any amino acid except proline.

The catalytic subunit of the mammalian OST (STT3) is thought to act during the co-translational insertion/translocation of proteins across the translocon. The OST is in fact integrated in the translocon and locates the STT3 catalytic site in the ER luminal face.

Briefly, during translocation, the nascent chain is directed towards OST catalytic site. In the ER luminal face, the catalytic subunit screens the nascent polypeptide looking for glycosylation acceptor sites (Figure 14). To achieve glycosylation the minimum distance between the end of a TMD placed at the Sec61 lateral gate and the catalytic site of OST (critical space) is about 6.5 nm (Braunger et al. 2018). This distance explains previous results from von Heijne and colleagues where they found that the distance, in amino acid residues, from the luminal end of a TMD and the glycosylation acceptor site is 14-15 residues if the acceptor site is placed upstream the TMD and 12-13 residues for an acceptor site placed downstream the TMD (Nilsson & G. von Heijne 1993).



**Figure 14 | Translocon components and scheme for co-translational N-glycosylation.** Schematic representation of ribosome-bound translocon complex with an interpolated example path for a nascent secretory protein. The STT3A catalytic site and a signal sequence (SP) or TMD in the Sec61 gate are separated by  $\sim 6.5$  nm. Translocon components represented in the scheme are colored: OST (red), TRAP (brown), and Sec61 (blue). Ribosome is dark gray colored. Taken from (Braunger et al. 2018).

## Introduction

### *i.2.3.3. Accessory translocon proteins*

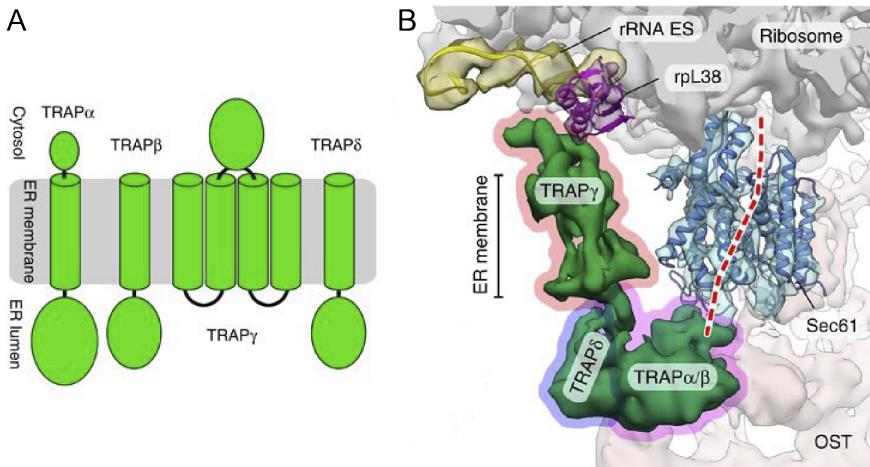
Some other membrane proteins as translocon-associated protein (TRAP), translocating chain-associating membrane protein (TRAM), PAT-10, RAMP4, or BAP31 have been reported to interact with the translocon and modulate its function at some stage.

TRAP is a tetrameric complex ( $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  subunits) associated with the translocon complex with a 1:1 stoichiometry (Ménétret et al. 2008) (Figure 15A). Despite the recent advances in the knowledge of the molecular organization of the TRAP complex, the precise function of this component is unknown.

The TRAP  $\alpha$  and  $\beta$  subunits form a hetero-dimer. The ER-luminal domain of the TRAP  $\alpha/\beta$  hetero-dimer contacts the loop of the hinge region between the N- and C-terminal halves of Sec61 $\alpha$ . This interaction positions the TRAP  $\alpha/\beta$  hetero-dimer directly below the channel pore (Pfeffer et al. 2017) (Figure 15B). In this position, the TRAP  $\alpha/\beta$  could interact with translocating nascent chains and influence the conformational state of the channel. Several studies suggest that the TRAP  $\alpha/\beta$  hetero-dimer is responsible for the observed effects of TRAP complex on the topogenesis of membrane proteins and the initiation of protein transport (Sommer et al. 2013; Fons et al. 2003).

The TRAP $\delta$  subunit is located at the periphery of the complex, at the interface between TRAP and the OST. Congenital disorders of glycosylation upon loss of TRAP $\delta$  suggest that it plays a role in coordinating the functions between TRAP and OST in mammals (Sommer et al. 2013; Ng et al. 2015).

The TRAP $\gamma$  subunit is predicted to contain four TMDs and a large cytosolic domain (~100 amino acid residues) (Figure 15A). The cytosolic domain of TRAP $\gamma$  binds to the ribosome on the cytosol face of the ER membrane and coordinates the remaining TRAP subunits and the interaction with the ribosome and the other translocon components.



**Figure 15 | Structure and subunit composition of the mammalian TRAP complex.**

(A) TRAP subunit composition and membrane topology as predicted by bioinformatic analysis. (B) Approximate positions of TRAP $\gamma$  (red outline), TRAP $\delta$  (blue outline), and the TRAP $\alpha/\beta$  hetero-dimer (magenta outline). The dashed red line is the path of a nascent polypeptide through the Sec61 complex (blue). Taken from (Pfeffer et al. 2017).

TRAM is an integral MP with eight TMDs and both N- and C-terminus facing the cytosol (Tamborero et al. 2011). Although it was recognized almost three decades ago as an essential component for translocation and insertion into the membrane (Görllich & Rapoport 1993), its precise function remains unknown. Nevertheless, the role of TRAM in the translocation of secretory proteins has been restricted to the insertion of the SS into the membrane. It has been found that TRAM is required for the insertion of SSs with short hydrophobic sequences or with low overall hydrophobicity. TRAM has also been reported to crosslink with TMDs from viral MPs (Saurí et al. 2007; Martínez-Gil et al. 2010), and with hydrophobic domains containing charged residues (Meacock et al. 2002; Cross & High 2009). These observations, together with the fact that TRAM itself contains an unusually high number of charged residues within its TMDs, led to the idea that TRAM could act as a chaperone for the integration of nonoptimal TMD by providing a more favorable context (Tamborero et al. 2011).

PAT10, a 10 kDa protein, was discovered as a translocon-associated protein during a Sec61's partners lookup (Meacock et al. 2002). Current results indicate that PAT-10 is part of a larger assembly (the PAT complex) that is adjacent to TMDs in process of insertion by the translocon. Hedge and colleagues suggested that the PAT complex acts as an intramembrane chaperone that protects low hydrophobic TMDs during assembly to minimize misfolding of multi-spanning MPs (Chitwood & Hegde 2020).

## Introduction

RAMP4 is a TA membrane protein implicated in promoting correct integration/folding of some integral MPs by facilitating subsequent glycosylation (Yamaguchi et al. 1999). In translocating ribosome-bound translocon complexes, RAMP4 is recruited to the Sec61 complex before the TMD emerges from the ribosome exit tunnel. Hence, it has been postulated that it is the presence of a TMD within the ribosome that triggers RAMP4 recruitment (Pool 2009)

BAP31 has also been reported to interact with the translocon complex. BAP31 is a multi-spanning integral MP that participates in the identification of misfolded proteins at the ER and promotes their subsequent retrotranslocation to the cytoplasm. The finding that BAP31 interacts with both Sec61 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 different channel functions may be performed in association with different cellular components. That is, the Sec61 complex might be merely the common player in a wide variety of transient complexes, each one performing different but related functions within biological membranes.

### **I.3. Insertion of membrane proteins**

The insertion and folding of a MP are often described as separated processes in a simplified two-stage model (Popot et al. 1987; Popot & Engelman 1990). In this model, insertion of individual TMDs occurs first. Once inserted, helices are envisioned to interact with each other within the membrane and adopt tertiary structures. Although this model may apply for some proteins (Popot et al. 1987), it is too simplified for others, which require the assistance of specialized machinery to acquire their native structure, e.g., it is known that some TMD can stay close to the translocon until stable interactions can be formed with subsequent helices (Sadlish et al. 2005; Saurí et al. 2005; Pitonzo et al. 2009).

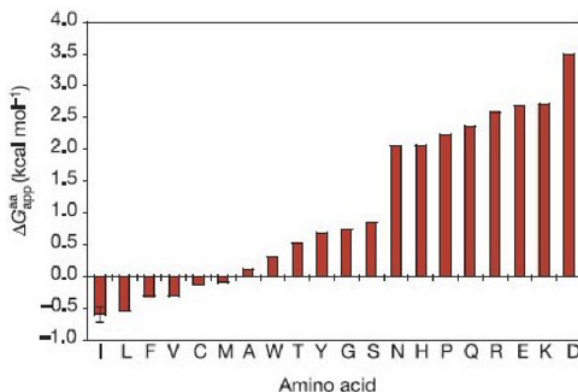
The insertion process depends on multiple factors, including hydrophobicity, amino acid composition, helical conformation, TMD orientation, and helix length and hydrophobic matching. Thus, it is not surprising that the membrane insertion process is considered a fine-tuned thermodynamic partitioning process where multitude of factors play an important role.

Several studies suggest that the translocon may act not only as a facilitator but also as a crucial component in the insertion/selection process, with a chaperoning function during the integration of optimal TMD into the membrane through favorable

acyl chain solvation (Johansson & Lindahl 2009). Indeed, mutations of the residues that form the hydrophobic constriction ring have been shown to influence insertion/translocation efficiency, modifying the hydrophobicity threshold for membrane insertion (Junne et al. 2010). Therefore, the translocon functioning may depend on some client nascent chain features discussed below.

### i.3.1. TMD hydrophobicity and amino acid preferences

Hydrophobicity is understood to be a measure of the relative tendency of a molecule to prefer a nonaqueous over an aqueous environment. Generally, the hydrophobicity of the TMD sequence drives integration into the membrane and it is determined by the side chain of the amino acids that form the TMD and the helical conformation, in which peptide bond polarity is significantly reduced (see below). To determine the hydrophobicity of peptide sequences, the partition coefficients of the natural amino acids to different solvents have been calculated and expressed in several hydrophobic scales (MacCallum & Tieleman 2011). These hydrophobic scales provide a starting point to recognize TMDs polypeptide sequences. A systematic study of amino acid sequences led to establish the efficiency of insertion of a TMD by the translocon (Hessa et al. 2005). These authors proposed a ‘biological’ hydrophobicity scale (Figure 16), in which the contribution from each of the 20 natural amino acids to the overall apparent free energy of membrane insertion ( $\Delta G_{app}^{aa}$ ) is given as a function of the position of the residue in the TMD (Hessa et al. 2007). This data not only take into account the importance of the amino acid side chain, but also the precise location of the naturally occurring 20 amino acids, which directly affect the helical conformation of the TMD.



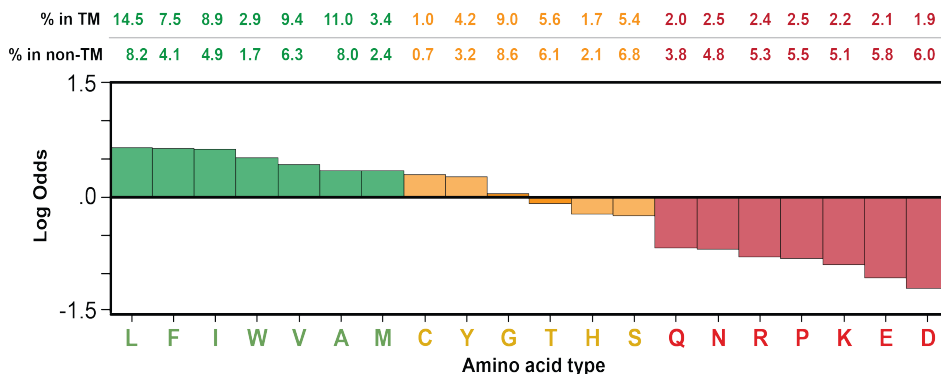
**Figure 16 | ‘Biological’ amino acid hydrophobicity scale.**

(A) Overall apparent free energy of membrane insertion ( $\Delta G_{app}$ ) scale derived from Helix-segments with the indicated amino acid placed in the middle of the 19-residue hydrophobic stretch. This scale establishes the efficiency of insertion by the translocon of the 20 naturally occurring amino acids placed in a TMD. Taken from (Hessa et al. 2005).

## Introduction

As expected the amino acid composition of  $\alpha$ -helical TMDs shows a preference for hydrophobic amino acids in comparison with loop regions (non-TMDs) in MP structures (Martínez-Gil et al. 2011; Baeza-Delgado et al. 2013) (Figure 17). Contrarily, charged amino acid residues do not show any preference to span biological membranes.

Surprisingly, the prevalence of Gly residues is almost equal in TMDs and non-TMDs regions of MP structures. Gly residues represent 9% of all amino acids found in TMDs, and have been associated with TMD packing (e.g., Glycophorin A dimerization) as they facilitate helix interactions due to its small side chain (Javadpour et al. 1999).



**Figure 17 | TMD amino acid preference.**

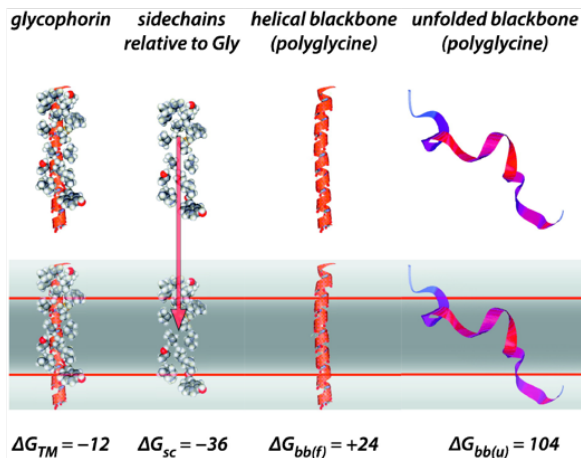
Amino acid preferences in TMD vs loop regions (non-TMD) in membrane protein structures. The top two rows show the percentage of occurrence of all amino acid types in TMDs and non-TMDs in membrane proteins of known structure. The lower plot shows the log odds of the occurrence. Positive log-odds indicate overoccurrence of the amino acid type in TMDs. Negative log-odds indicate the underrepresentation of the amino acid type in TMDs. Amino acids are colored according to an arbitrary division of their log-odds (i.e. green for log-odds > 0.3; orange for  $0.3 \leq \log\text{-odds} < 0.3$ ; and red for log-odds < -0.3). Taken from (Martínez-Gil et al. 2011).

### i.3.2. Helical conformation of TMD

The hydrophobic effect between the residues and the membrane depends upon complete dehydration of the non-polar structure surface when the polypeptide chain is moved from the bulk water to a non-polar phase, i.e. the transition from the ribosome-translocon tunnel to the lipidic bilayer. To address this hydrophobic effect the  $\alpha$ -helix formation is a critical key point.

The  $\alpha$ -helix structure formation allows, via maximizing hydrogen bond formation, a dramatic reduction of the unfavorable free energy cost of partitioning non-hydrogen-bonded peptide bonds. This reduction can be noticeable when

comparing the energetic cost of insert a unfolded polyglycine backbone ( $\Delta G_{bb(u)}$ ) and the same but helical folded backbone ( $\Delta G_{bb(f)}$ ) (Figure 18). However, it should be noted that hydrogen bond formation during  $\alpha$ -helix structure acquisition is not enough to drive a thermodynamically favorable partitioning into the lipid bilayer. The thermodynamic cost of dehydrating peptide bonds upon insertion has been estimated to be about +1.2 kcal/mol per residue (Jayasinghe et al. 2001), which remains non-favorable (Figure 18,  $\Delta G_{bb(f)}$ ). This phenomenon explains why amino acid residues that compose TMDs must be very hydrophobic to compensate for the unfavorable cost of partitioning in helical conformation by the favorable cost of partitioning hydrophobic side chains (Figure 18, left).



**Figure 18 | Amino acid helix conformation.**

The energetics of inserting an  $\alpha$ -helix into lipid bilayers (gray) is illustrated here for the Glycophorin A TMD helix. Free energy of the TMD when is folded ( $\Delta G_{TM}$ ). Free energy relative only to Glycine side chains ( $\Delta G_{sc}$ ), determined by the hydrophobic effect, must compensate for the unfavorable folded backbone  $\Delta G_{bb(f)}$ .  $\Delta G_{bb(u)}$  is the energetic cost for the unfolded polyglycine backbone immerse in the bilayer. Taken from (Cymer et al. 2015).

### i.3.3. Hydrophobic matching

As previously mentioned, a ‘hydrophobic mismatch’ occurs when the hydrophobic length of a TMD does not match the hydrophobic thickness of the membrane core. Consequently, two types of hydrophobic mismatch could arise: Positive mismatch, when the membrane thickness is shorter than the TMD length; and negative mismatch, when the length of the hydrophobic section of a TMD is shorter than the hydrophobic core of the membrane (Figure 19).

In both scenarios, either the membrane or the polypeptide should adapt to minimize the exposure of hydrophobic residues to the aqueous media (positive mismatch) or the extrusion of polar amino acids within the hydrophobic core of the membrane (negative mismatch) (Andersen & Koeppe 2007). Both rearrangements are known to be important for determining the final assembly of a MP as shown by

## Introduction

fluorescence (Ren et al. 1999; Sparr et al. 2005) and chimeric overexpression of dimerizing TMD in membrane-mimetic environments (Orzáez et al. 2000; Orzáez et al. 2005).

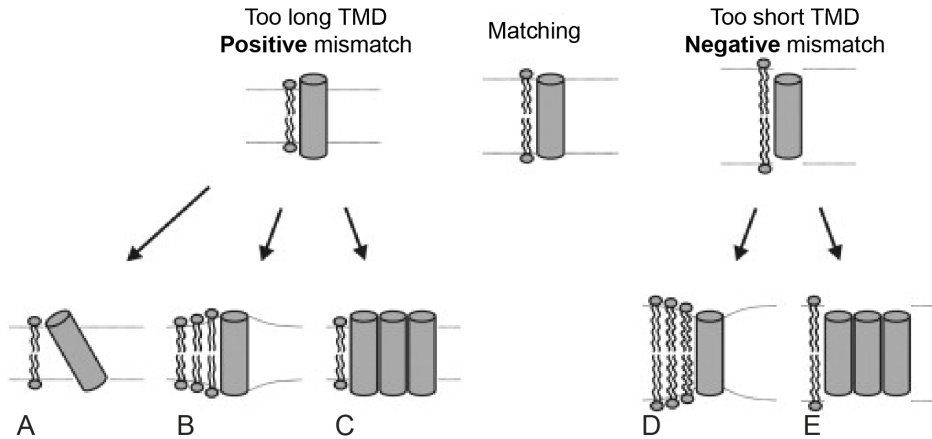
For **positive** hydrophobic mismatch one or more of the following adaptations can occur:

1. The  $\alpha$ -helix can reduce its hydrophobic length by adopting a more compacted conformation ( $\pi$ -helix).
2. The peptide can tilt, reducing its exposure to polar groups outside the membrane (Figure 19A).
3. The acyl chains near the peptide can arrange, increasing the local bilayer hydrophobic width by moving closer to the TMDs phospholipids with longer acyl chains (Figure 19B).
4. The peptides can oligomerize or aggregate, thus reducing the exposure of hydrophobic groups to the bulk water outside the membrane (Figure 19C).
5. The peptide can be expelled from the bilayer, so it will look for a thicker bilayer in other organelles.

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

1. The  $\alpha$ -helix can increase its hydrophobic length by adopting a more stretched conformation (eg.  $3_{10}$  helix).
2. The bilayer width near the peptide can decrease, by acyl chain disordering (Figure 19D).
3. The peptides may aggregate or oligomerize, thus reducing the exposure of polar groups to the membrane hydrophobic core (Figure 19E).
4. The peptide can induce non-lamellar phase formation, which results in a membrane thickness reduction that can match with the peptide hydrophobic length.
5. The peptide can be expelled from the bilayer, so the peptide will look for a thinner bilayer in other organelles.





**Figure 19 | Some of the possible adaptations to hydrophobic mismatch.**

In the case of too-long TMD peptides (positive mismatch): (A) peptide tilting, (B) bilayer distortion, and/or (C) peptide aggregation. For too-short TMD peptides (negative mismatch): (D) bilayer distortion and/or (E) peptide aggregation. Taken from (Ramadurai et al. 2010).

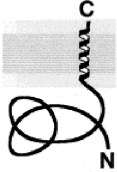
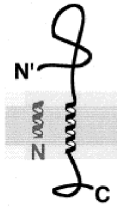
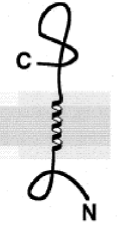
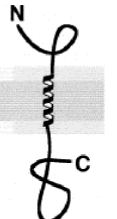
## I.4. Topology of membrane proteins

### i.4.1. Topology of single-spanning membrane proteins

Membrane proteins can be classified based on their topology, that is, the orientation that they acquire in the membrane.

Single-spanning MPs, which span the bilayer only once, can only assume a cytoplasmic N-t and an exoplasmic C-t ( $N_{\text{cyt}}/C_{\text{exo}}$ ) orientation or the reverse  $N_{\text{exo}}/C_{\text{cyt}}$  topology. However, when the mechanism of insertion is taken into consideration, four major types of single-spanning MPs can be distinguished, as summarized in Figure 20.

## Introduction

Signal type:	C-Terminus-translocating signals			N-Terminus-translocating
Topogenic determinants:	C-terminal signal	Cleaved signal + stop transfer sequence	Signal-anchor SAII	Reverse signal-anchor SAI
Machinery:	TRC40-Get3 pathway	SRP/SR/Sec61 + signal peptidase	SRP/SR/Sec61	SRP/SR/Sec61
exoplasmic	 <p style="text-align: center;">Type IV</p>	 <p style="text-align: center;">Type I</p>	 <p style="text-align: center;">Type II</p>	 <p style="text-align: center;">Type III</p>
Final topology:				
Examples:	Synaptobrevin; cytochrome b5	Glycophorin; LDL receptor	Transferrin receptor; galactosyl transferase	Synaptotagmin; neuregulin; cytochromes P-450

**Figure 20 | Types of topogenic determinants of single-spanning membrane proteins.**

Single-spanning membrane proteins are classified based on their membrane topology. The machinery involved in their insertion and some examples of each type are included. Adapted from (Goder & Spiess 2001).

Type I MPs are targeted to the ER membrane by a cleavable N-t SS, typically composed by 7-15 non-polar amino acids flanked by positively charged residues at the N-t. The SS itself generally adopts an  $N_{\text{cyt}}/C_{\text{exo}}$  orientation. Next, the SS is cleaved by the signal peptidase. The presence of a downstream TMD will stop translocation of the following polypeptide region across the ER membrane and set a final  $N_{\text{exo}}/C_{\text{cyt}}$  protein topology. Accordingly, the TMD in Type I MP is called anchor domain or stop-transfer sequence because it stops the translocation and attaches the protein into the membrane.

When there is no SS, single-spanning MPs can adopt an  $N_{\text{cyt}}/C_{\text{exo}}$  (Type II) or  $N_{\text{exo}}/C_{\text{cyt}}$  (Type III) topology. In these cases, the TMD is responsible for both, targeting and anchoring the protein into the membrane. In these proteins the final adopted topology will depend on the so-called ‘topological determinants’, explained in i.4.3.

There is one more type of single-pass MPs known as Type IV or TA MPs. In these proteins the TMD is located near the C-terminus, forcing a final  $N_{\text{cyt}}/C_{\text{exo}}$  topology. In contrast with Type I, II, and III which are inserted by the SRP-SRP receptor and Sec61 machinery, Type IV MPs are inserted post-transcriptionally by the TRC40-Get pathway (see i.2.2.1), since the signal to target the protein to the membrane emerges from the ribosome once protein translation has been completed.

### i.4.2. Topology of multi-spanning membrane proteins

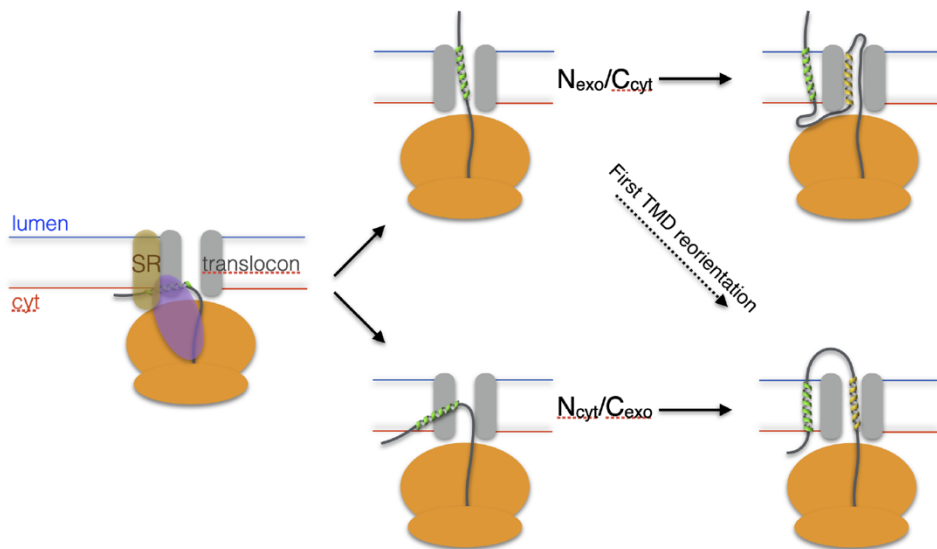
For multi-spanning MPs, the integration of the first TMD into the ER membrane is considered key to define the protein topology. This first TMD can either be inserted with its N-t in the ER lumen or the N-t facing the cytoplasm (Figure 21).

The insertion of proteins with the first TMD with a  $N_{\text{exo}}/C_{\text{cyt}}$  topology may be considered as a head-first insertion, give rise to a protein with the N-t translocated followed by a TMD acting as a signal anchor. After SRP recognition of the TMD it is easy to imagine a short amino acid sequence placed before the first TMD being translocated through the translocon. Nevertheless, it is less clear how longer N-t sequences can be pushed through the translocon by the same mechanism. For long N-t sequences, it seems that they should remain partially unfolded to be translocated (Kida et al. 2005), being the translocation inhibited if rapidly folding hydrophilic domains are placed at the N-terminus of the protein (Denzer et al. 1995). After the first TMD the insertion of the second TMD requires a hairpin rearrangement that is a TMD flipping (Figure 21, top)

Proteins with the first TMD with a  $N_{\text{cyt}}/C_{\text{exo}}$  sequences translocate their C-t residues into the ER lumen and keep the N-terminus on the cytoplasmic side of the membrane. To acquire this orientation, the TMD has to go through a perpendicular disposition to the translocon (flipping process) (Figure 21, bottom). In this scenario, once the topology of the first TMD has been fixed the insertion of the second TMD may be a simple head-first insertion.

Despite being widely considered the orientation of the first TMD as the main protein topology determinant, there is good evidence that in some cases an initial topology can be reversed (Goder & Spiess 2003). These ‘gymnastics’ have been demonstrated in the assembly of aquaporin-4 where its first TMD initially inserts with an  $N_{\text{out}}$  orientation before reversing to its final  $N_{\text{cyt}}/C_{\text{exo}}$  orientation as the nascent chain elongates (Devaraneni et al. 2011). These findings corroborate that flipping and final protein topology could be influenced by nascent polypeptide length, charge difference, and signal sequence hydrophobicity (Goder & Spiess 2003).

## Introduction



**Figure 21 | Routes of possible integration of the first two TMDs of membrane proteins.**

Once the SRP (purple colored) dissociates from the SE and the ribosome transfers the hydrophobic signal to the translocon and elongation of the nascent polypeptide resumes. The insertion of the first TMD (green helix) may be a simple head-first process ( $N_{\text{exo}}/C_{\text{cyt}}$ ) or may require a hairpin rearrangement ( $N_{\text{cyt}}/C_{\text{exo}}$ ). A second hydrophobic region (top panel, depicted in yellow) downstream of a  $N_{\text{exo}}/C_{\text{cyt}}$  sequence would require helix flipping, while its location downstream of a sequence with  $N_{\text{cyt}}/C_{\text{exo}}$  would facilitate a head-first helix insertion (bottom). The dotted line indicates that the initial insertion of the first TMD can be reversed (dotted arrow). Adapted from (Whitley & Mingarro 2014).

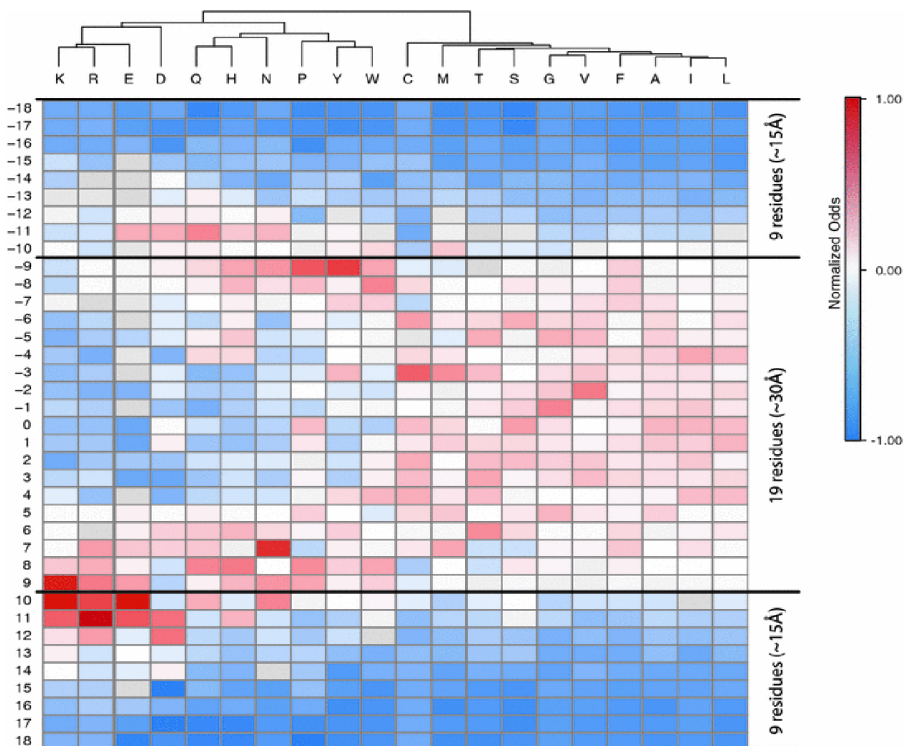
### i.4.3. Orientation of transmembrane helices

During insertion, nascent membrane proteins have to adopt the correct topology in the lipid bilayer to achieve their native structure. Whether a TMD adopts an N-terminal cytosolic ( $N_{\text{cyt}}$ ) or a N-terminal luminal ( $N_{\text{lum}}$ ) orientation depends on several factors.

The distribution of charged residues between both flanking regions of a TMD is a major topology determinant in MPs (G. von Heijne 1986; G. von Heijne 1989) (Figure 22). The so-called ‘positive-inside rule’ was firstly observed in bacteria cells. It was noted that the net negative-inside electrical potential across the membrane together with the enrichment in negatively charged lipids in the cytoplasmic bilayer leaflet could be responsible for TMD orientation. Later on, a similar skewed amino acid distribution was also identified in eukaryotes (Hartmann et al. 1989; Goder & Spiess 2001). Moreover, it has been demonstrated that certain residues of the

translocon also contribute to the positive-inside orientation of signal sequences (Junne et al. 2007; Goder et al. 2004).

Multi-spanning MPs generally adopt their native orientation depending on the insertion of the signal sequence or the first TMD, which determines the orientation of subsequent TMDs. Nevertheless, it has been demonstrated that the topology of a full-length multi-spanning protein can be controlled by a single C-terminal residue (Seppälä et al. 2010). Experimental evidence has been found for some proteins that multiple topologies are possible depending on the cellular localization or environment (Hegde et al. 1998), highlighting again the importance of the bilayer. On the other hand, other multi-spanning MPs, such as the viral MP p9, have a strong preference for a specific topology determined by multiple, perhaps redundant, sequence features implied in the maintenance of the TMD orientation (Saurí et al. 2009).

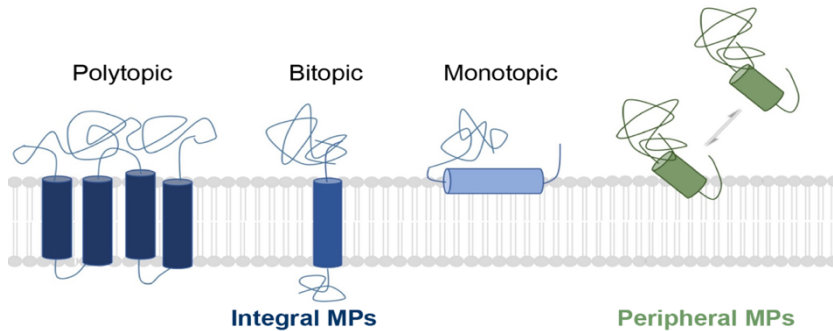


**Figure 22 | Amino acid type and position distribution in TMD helices.**

Each amino acid type and its positioning in the TMD 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 based on 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.4.4. Monotopic membrane proteins

Based on their topology, MPs can also be described as monolayer-integrated proteins or monotopic MPs (MMPs). A MMP is a MP that does not span the membrane bilayer but rather associates to a single face on the membrane bilayer (K. N. Allen et al. 2019) (Figure 23).



**Figure 23 | Structural representation of membrane proteins**

Integral MPs are classified as polytopic (or multi-spanning, that spans the membrane multiple times), bitopic (or single-spanning, that spans the membrane a single time), or monotopic (that does not span the bilayer but remains attached on a single face of the membrane) dependent on their membrane-associated topology and hydrophobic segment distribution across the membrane. Peripheral MPs associate reversibly with the lipid bilayer (right). Taken from (K. N. Allen et al. 2019).

How the hydrophobic membrane domains of MMPs interact with membranes has not been systematically investigated, primarily because their sequence composition are difficult to distinguish from the TMDs of MPs, but also because the fluid nature of the lipid bilayer makes experimental studies extremely challenging. In the last years, efforts have been focused to provide a throughput method to identify and study MMPs. However, currently fewer than 100 MMPs have been unambiguously identified ([blanco.biomol.uci.edu/mpstruc/](http://blanco.biomol.uci.edu/mpstruc/)) and our knowledge about how they are targeted to and interact with membranes is far from being complete.

Among MMP the most common membrane association is that of amphipathic helices (AHs) positioned parallel membrane interface (A. L. Lomize et al. 2007; Pataki et al. 2018). AHs are protein sequences that fold into a helical structure upon contact with a polar/non-polar interface, as the surface of a biological membrane (Cornell & Taneva 2006). In such sequences, hydrophobic amino acids are regularly distributed with polar residues in between, thereby allowing the helix to present two faces with opposite chemical features: a hydrophobic face and a hydrophilic face. Except for their amphipathic character, the AHs are usually different in their composition (length and amino acid sequence) and in their surface-binding

properties: length ranges between 12 and 950 amino acid residues; the hydrophobic amino acids implied are usually different; the charged residues in an AH can be positively or negatively charged, or a combination of both, not following a characteristic pattern (Giménez-Andrés et al. 2018).

Examination of the limited MMPs structures available also revealed other common features present in MMPs as hydrophobic loops extending into one leaflet of the membrane, or functional dimmers that increase the surface area of the protein-membrane interaction (G. von Heijne 2006; Mattevi 2006; Balali-Mood et al. 2009).

#### i.4.5. Packing motifs of membrane proteins

In recent years, the complete topology of membrane-embedded  $\alpha$ -helical MPs has been discovered to be in part determined by the interaction between neighbor helices. These helices usually pack through van der Waals interactions. In most cases these forces are maximized in some proteins through a limited number of sequence motifs that minimize the interhelical distance through sidechain packing (Liljas et al. 2009). These packing motifs differ from those found in the TMDs of channels and transporters where helix-helix interfaces are not tightly packed as they have to leave room for the passage of crossing molecules.

A bioinformatic analysis (Walters & DeGrado 2006) revealed that the most commonly distributed packing motif is the Ala-coil-like motif. This motif was also termed GAS, to reflect the frequency residues Gly, Ala, or Ser. In the GAS motif 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 proximity. Between the motifs present in proteins, the perfect GxxxGxxxG sequence (also known as glycine zipper-like motif) has been found to occur in more than 7% of all known MPs structures. Furthermore, approximately 25% of all MPs contain one of the non-perfect but strong glycine zipper-like motifs (G,A,S)xxxGxxxG and GxxxGxxx(G,S,T) (S. Kim et al. 2005). Those motifs, depending on sequence context, can induce dimerization of TMD helices (e.g., as the widely studied glycoporphin A (GpA) TMD homo-dimer (GVxxGVxxT) (Langosch et al. 1996; Smith et al. 2001; Orzáez et al. 2005)).

Minimized variants of the glycine zipper-like motifs can be found in a wide range of TMDs. These motifs only present two small residues (Gly, Ala, or Ser) separated by three amino acids in the polypeptide chain (small-xxx-small). The co-location of these two small residues in the polypeptide chain results in a groove in the helix which allows two helices to interlock, maximizing van der Waals contacts.





# **OBJECTIVES**



The main objective of the present thesis is to increase our knowledge on membrane protein biogenesis and folding in the biological context of the translocon. During the thesis the following specific objectives have been addressed:

- Folding of helices within the ribosomal exit tunnel.
- Development of a novel assay to determine the interfacial disposition of protein segments in a quantitative manner.
- Establishment of an interfacial scale for the 20 naturally occurring amino acid residues.
- Study of the hydrophobic effect on TMD-TMD packing in eukaryotic membranes.
- Topology determination of the subunit gamma from the TRAP complex.
- Topogenesis of bacteriorhodopsin in microsomal membranes.



# **MATERIALS AND METHODS**



## ***In vitro* transcription and translation**

### Plasmid preparation

To subclone TMD sequences into the ‘Lep’ system the corresponding DNA was either PCR amplified, synthesized by Invitrogen (*GeneArt* gene synthesis), or constructed by oligonucleotides annealing. Next, the DNA segments were subcloned using *SpeI* and *KpnI* restriction sites into *Lep* plasmid (pGEM1-*Lep*, provided by Dr. von Heijne (Hessa et al. 2005)).

The DNA encoding full-length proteins was synthesized by Invitrogen (*GeneArt* gene synthesis) and subcloned into pGEM1 using the In-Fusion HD technology (Takara), according to the manufacturer’s instructions.

### *Site-directed mutagenesis*

Point mutations were designed and carried out using the Quik Change II Primer Design tool and the PfuTurbo DNA polymerase (Agilent Technologies), following the manufacturer’s instructions. Briefly, a 50  $\mu$ L mutagenesis reaction mixture included: 50 ng of template DNA, 5  $\mu$ L of 10X reaction buffer, 1  $\mu$ L of the PfuTurbo DNA polymerase, and 0.05  $\mu$ M (final concentration) of each primer. All DNA manipulations were confirmed by DNA sequencing (Macrogen).

### *In vitro* protein expression

DNA templates were amplified using a forward primer including the T7 promoter and a reverse primer with or without stop codon, according to the needs of the assay.

PCR products were transcribed and translated using TNT T7 Quick Coupled System (Promega). The reaction contained: 5  $\mu$ L of TNT T7, 1  $\mu$ L of the direct PCR product mixture, 0.5  $\mu$ L of EasyTag EXPRESS <sup>35</sup>S Protein Labeling Mix (5.5  $\mu$ Ci, Perkin Elmer), and 0.3  $\mu$ L of ER rough microsomes (tRNA Probes). Translation mixtures were incubated at 30°C for 30 or 90 min, depending on the absence or presence of stop codon, respectively. Reactions were stopped by adding 50  $\mu$ L of 1x SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and visualized by micro-autoradiography using a Fuji FLA-3000 phosphoimager. Bands were quantified using ImageJ (NIH).

## Materials and methods

### Endoglycosidase H and Proteinase K treatments

To perform the Endoglycosidase H assay 18  $\mu\text{L}$  of the *in vitro* translation mixture with 1x SDS-PAGE sample buffer were mixed with 2  $\mu\text{L}$  of the NEB GlycoBuffer 3 and 2 U of Endoglycosidase H enzyme (New England Biolabs) and incubated for 30 min at 37°C. After incubation, products were analyzed by SDS-PAGE and autoradiography.

To perform the proteinase K protection assay, 5  $\mu\text{L}$  of the translation mixture (before the addition of SDS-PAGE sample buffer) were supplemented with 1  $\mu\text{L}$  of proteinase K (2 mg/mL, Merck), and incubated 20 min on ice. The reaction was stopped by adding 120  $\mu\text{L}$  of 200  $\mu\text{M}$  phenylmethanesulfonyl fluoride (PMSF, Merck) in PBS. Next, membranes were collected by ultracentrifugation (100,000 x g, 4°C 15 min) on a 0.5 M sucrose cushion. Supernatants were carefully removed and pelleted microsomes were resuspended in 40  $\mu\text{L}$  of 1X SDS-PAGE sample buffer. Samples were analyzed by SDS-PAGE and autoradiography.

## Confocal microscopy

### Immunofluorescence

Ten mm diameter glass cover-slides were treated with 50  $\mu\text{L}$  of Poly-L-Lysine 0.01% (Merck) and washed with PBS (3x). HEK-293T cells were cultured on the glass slides ( $4 \times 10^4$  cells/plate in 24 wells plate). After 24 hours, cells were transfected using Polyethyleneimine (PEI, Merck) as previously described (Longo et al. 2013).

After 48 hours, cells were fixed with 4% formaldehyde (Merck) and washed in PBS (3x). Cell permeabilization was done with a solution of PBS, 1% BSA, 0.1% Triton X-100 for 2 minutes. Immuno-stainings were done using a primary  $\alpha$ -Flag M2 antibody (Sigma), followed by a secondary Alexa647-conjugated anti-mouse antibody (Sigma). Additionally, cells were DAPI stained (Fluoroshield, Merck). Confocal images were captured at the Microscopy core facility of the SCSIE (University of Valencia) on FV1000 confocal microscope (Olympus) using a 60x oil lens.



## Dimerization assays

### ToxRed assay

The DNA coding for the tested TMDs (Table 1) was introduced into the ToxRed plasmid (provided by Dr. Berger (P.-C. Su & Berger 2013)) using the *XhoI* and *HindIII* restriction sites (Figure 49A). Note that, selected restriction sites incorporated a Leu and a Lys residue at the N-terminus of the chimeric TMD. All sequences were corroborated by DNA sequencing (Macrogen).

The ToxRed assay was performed using the malE- *E. coli* (MM39) strain. *E. coli* MM39 cells were transformed by electroporation with the appropriate plasmids and plated onto LB with 100 µg/mL ampicillin (LBA) plates. After o/n incubation at 37°C, individual colonies were selected, inoculated into LBA medium, and grown at 37°C o/n. Next, cell bacteria cultures were diluted (1/100) in M9 minimal medium supplemented with 0.8% maltose and grown at 37 °C o/n. Finally, the red fluorescence (595nm) and optical density (620 nm) were measured in a Multimode Plate Reader Victor X3 (Perkin Elmer).

Western blots of the transformed proteins were done as follows. *E. coli* MM39 colonies were inoculated into LBA medium and grown at 37 °C (OD<sub>420</sub> of 0.6). Next, cultures were harvested by centrifugation, resuspended in lysis buffer (TBS (Tris-HCl 20 mM pH 7.5, NaCl 150mM), 1% NP-40) and freeze-thaw 3 times. The suspensions were clarified by centrifugation (13,000 x g). The supernatants were mixed with SDS-PAGE sample buffer, heated 5 min to 95°C, and loaded on 12% SDS-polyacrylamide gel. Next, proteins were transferred onto nitrocellulose membranes. The constructs of interest were detected with HRP-conjugated α-MBP monoclonal antibody (New England Biolabs) and ECL reagent (GE Healthcare). Chemiluminescence was visualized by and ImageQuant LAS 4000 (GE Healthcare).

### BIFC

The TMD sequences were PCR amplified or synthesized (Invitrogen, *GeneArt* gene synthesis) and subcloned into modified BIFC plasmids (pBIFC-VN172 [VN] and pBIFC-VC55 [VC], provided by Dr. Orzáez (Andreu-Fernández et al. 2017)). TMDs were placed at the C-terminus of the VN or VC half of the Venus Fluorescent Protein (VFP) using *NotI* restriction enzyme (Figure 50). For EphA2-derived sequences, the TMD sequences were PCR amplified from the full-length sequence and subcloned into BIFC plasmids at the N-terminus of the VN or VC half of the VFP using *EcoRI* and *XhoI* restrictions sites. All sequences were corroborated by DNA sequencing (Macrogen).

## Materials and methods

The BiFC assay was performed in HEK-293T cells as follows. Cells were grown in Dulbecco's modified Eagle medium (Gibco) supplemented with 10% fetal bovine serum. Cells were plated in 24-well plates at 37°C, 5% CO<sub>2</sub>, at a concentration of 2x10<sup>6</sup> HEK-293T cells/plate and transfected with plasmids encoding VN and VC (250 ng each) using PEI (Merck), as previously described (Longo et al. 2013). Twenty-five ng of a plasmid encoding the Renilla luciferase under the CMV promoter (pRL-CMV, Promega) were co-transfected for normalization purposes. After 48 hours luciferase and fluorescence were measured using a Multimode Plate Reader Victor X3 (Perkin Elmer). The luciferase signal was measured using the Renilla Luciferase Flash Assay Kit (Thermo Scientific), according to the manufacturer's instructions. Immuno-identification of the VN and CN chimeras were done using  $\alpha$ -c-Myc and  $\alpha$ -HA rabbit antibodies, followed by a secondary HRP-conjugated  $\alpha$ -rabbit antibody (Merck). Chemiluminescence was visualized by an ImageQuant LAS 4000 (GE Healthcare).

Confocal fluorescence microscopy of the BIFC plasmids was done as mentioned above. Cells were co-transfected with 200 ng of a plasmid encoding an ER marker (mCh-Sec61 $\beta$ , Addgene 49155) or a PM marker (mCh-Mem, Addgene 55779) to facilitate sub-cellular localization.

## Subcellular fractionation procedure

After transfection cells were resuspended in 500  $\mu$ L of subcellular fractionation buffer (250 mM Sucrose, 20 mM HEPES pH 7.4, 10 mM KCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EDTA, 1mM EGTA) and lysed by sonication. Cell lysates were incubated at 4°C for 30 min and centrifuged (10,000 x g, 4°C, 10 min). Supernatants were carefully transferred to a new tube and ultracentrifuged (100,000 x g, 4°C 1h). The supernatant (cytosolic fraction) was stored at -20 °C while the pellet was re-suspended in 500  $\mu$ L of subcellular fractionation buffer and ultracentrifuged (100,000 x g, 4°C 1h). The pellet (membrane fraction) was re-suspended in 150  $\mu$ L of 1x SDS-PAGE sample buffer.

## Peptide characterization

Peptides were chemically synthesized by Peptide 2.0 Inc. (Chantilly, VA) using Fmoc chemistry. Peptide purification was done by HPLC and assessed by MALDI-TOF. BaxE5 peptide stocks were prepared using MiliQ water and pH was adjusted to 8 with NaOH. Peptide stock concentration was calculated by absorbance of the single Tyr residue in the sequence, using a molar extinction coefficient of  $1490 \text{ M}^{-1} \text{ cm}^{-1}$ . Final stock concentration was adjusted to  $10 \mu\text{M}$ .

## Circular dichroism

POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, Avanti polar lipids), a frequently used phosphatidylcholine lipid, was dried in a dark vacuum chamber, rehydrated in  $1 \text{ mM}$  NaPi buffer pH 7.5, and extruded in  $100 \text{ nm}$  pore size filters, resulting in large unilamellar vesicles (LUV).

Peptide working solution was prepared by diluting the  $10 \text{ mM}$  peptide stock in  $1 \text{ mM}$  phosphate buffer pH 7.5. Peptides were incubated with  $0.1 \mu\text{m}$  diameter POPC LUVs to reach a peptide:lipid molar ratio of 1:200 in  $1 \text{ mM}$  NaPi buffer (pH 7.5) for at least 45 min. After peptide-lipid incubation, pH was modified adding different pH buffers (sodium phosphate or sodium acetate buffer) to obtain a  $5 \text{ mM}$  final buffer concentration and a  $5 \mu\text{M}$  final peptide concentration. Samples were incubated for an additional 45 minutes. CD measurements were performed on a Jasco J-815 spectropolarimeter at  $25 \text{ }^\circ\text{C}$  in a  $2 \text{ mm}$  cuvette. Data were collected every  $1 \text{ nm}$  at a speed of  $50 \text{ nm/min}$  in the range of  $260$  to  $195 \text{ nm}$ , with a bandwidth of  $2 \text{ nm}$  and 20 accumulated scans. The signal from the lipid backgrounds was subtracted from the data. Resulting values were normalized to ellipticity milligrades using the equation:

$$\Theta \text{ (deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \cdot 10^3) = \frac{\text{ellipticity (mdeg)} \cdot 10^6}{\text{cuvette length (mm)} \cdot [\text{peptide}] (\mu\text{M}) \cdot n}$$

where  $n$  is the number of amino acids in the peptide minus 1 (number of peptide bonds).

## Sulforhodamine B leakage assay

Dried POPC lipids were rehydrated with  $20 \text{ mM}$  Sulforhodamine B (S1307, Thermo) in water and extruded in  $200 \text{ nm}$  filters. Excess of free Sulforhodamine B was removed using a Sephadex G-25 desalting column (GE Healthcare). The elution of sulforhodamine B encapsulated POPC vesicles was done using  $1 \text{ mM}$  NaPi buffer

## Materials and methods

(pH 7.5). To perform the assay the peptide was freshly diluted from the 10  $\mu\text{M}$  stock in 1 mM NaPi (pH 7.5). Finally, the peptide and lipid mixture was buffered with 150 mM buffers according to the pH needs (sodium phosphate or sodium acetate buffer, as above).

The peptide:lipid molar ratios in the leakage assay were modified by keeping vesicle concentration constant (90  $\mu\text{M}$ ) and changing the peptide concentration from 3 to 500 nM in 200  $\mu\text{L}$  reaction volume. The assay was performed using a BiotechCytation 5 Imaging Reader (Biotech) to measure the increase of fluorescence (485/590 nm) caused by leakage. The percentage of leakage was calculated using the equation:

$\% \text{ leakage} = \frac{\text{sample-control}}{\text{triton-control}} \times 100$ , where Triton X-100 was used as total leakage control, and non-addition of peptide or Triton X-100 was used as non-leakage control.

# RESULTS



## Chapter 1. Early stages of membrane protein folding

Membrane-spanning domains of integral MPs must achieve their final folded structure in a very different environment, the hydrophobic interior of a lipid bilayer, compared to that experienced by soluble proteins. As explained in section 1.3.2, in the membrane environment there is a strong driving force for polypeptide chains to adopt regular secondary structures (mainly  $\alpha$ -helical) to reduce the significant free energy penalty of exposing polar peptide bonds to the hydrophobic core of biological membranes. Thus, the formation of  $\alpha$ -helices, stabilized by a regular hydrogen bond network of polar peptide bonds, is essential for the folding and insertion of TMD into biological membranes.

In the biogenesis of all proteins, the nascent polypeptide must navigate through the ribosomal tunnel toward the exit site. For the vast majority of eukaryotic integral MPs, nascent chains are elongated by ribosomes following targeting of a translationally stalled RNC-SRP complex to the translocation/insertion machinery, i.e., the Sec61 translocon in the ER membrane. The translocon facilitates the insertion of TMDs into the lipid bilayer (Johnson & van Waes 1999) in addition to the translocation of the luminal region of MPs and secreted proteins across the ER membrane (Rapoport 2007; Whitley & Mingarro 2014). The alignment of the ribosome exit site with the central pore of the translocon is proposed to facilitate direct movement of the elongating polypeptide chain from the ribosomal exit tunnel across or into the membrane (Becker et al. 2009). The internal diameter of both the ribosomal exit tunnel (Ban et al. 2000) and the translocon (Van den Berg et al. 2004) range from  $\sim 10$  to  $20 \text{ \AA}$  (Gumbart et al. 2011; Voss et al. 2006), which have been shown to be sufficient to allow secondary structure formation of  $\alpha$ -helices in elongating nascent polypeptide chains (Mingarro et al. 2000; L. W. Tu & Deutsch 2010; Bhushan et al. 2010).

The importance of co-translational folding while the nascent polypeptide chain is still tethered at the ribosomes has been well-established (Zhang & Ignatova 2011). The folding of tethered nascent chains into  $\alpha$ -helical conformation in the ribosomal tunnel has also been demonstrated (Mingarro et al. 2000; Whitley et al. 1996; Woolhead et al. 2004; Lu & Deutsch 2005a; L. W. Tu & Deutsch 2010; Bhushan et al. 2010). Despite this, it remains unclear what features of a helical region influence the propensity to acquire an  $\alpha$ -helical structure whilst still in the ribosome. In particular, given that a TMD should be folded prior to its exposure to the lipidic environment for thermodynamic reasons (Popot & Engelman 2000; White & Wimley 1999), we considered that  $\alpha$ -helical TMD might achieve secondary structure in a different location/environment than helices from water-soluble proteins.

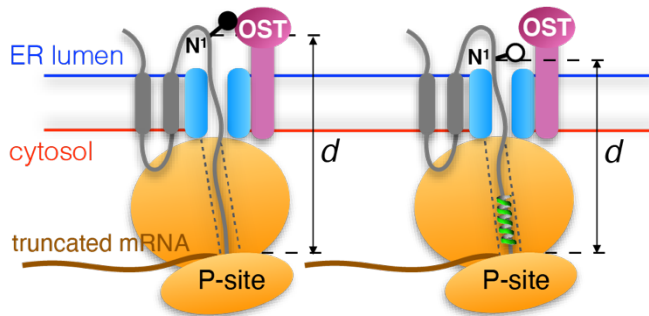
## Results

To address this possibility, truncated nascent chains trapped within the ribosome-translocon complex were used. The nascent chains designed encoded for the model protein Lep containing engineered ‘test’ sequences of amino acids with known helical conformation in the final folded protein. Whereas these test sequences had different biophysical properties, i.e. were hydrophobic TMD stretches of amino acids or hydrophilic non-TMD (soluble) sequences of similar length. We measured the number of residues of nascent polypeptide ( $d$ , distance P-NST) required to span the distance between the P-site on the ribosome (placed at the entrance to the ribosomal tunnel) and the active site of the OST (located nearby the luminal end of the translocon central pore) (Wild et al. 2018; Bai et al. 2018).

### c.1.1. The glycosylation mapping assay

For a trapped polypeptide within the ribosome-translocon complex, the number of residues required to bridge the distance between the ribosomal P-site and the active site of the OST can be conveniently measured by glycosylation mapping (Mingarro et al. 2000; L. W. Tu & Deutsch 2010). Radiolabeled, fully assembled translocation intermediates can be prepared *in vitro* by translating truncated mRNAs (lacking a 3' stop codon within the coding region) in the presence of  $^{35}\text{S}$ -labeled amino acids and dog pancreas microsomes. A ribosome halts when it reaches the end of such an mRNA, but the nascent chain remains tethered to the ribosomal P-site because the absence of a stop codon prevents normal termination from occurring (Figure 24). A series of neighboring truncation points on the mRNA are tested such that a unique Asn-Ser-Thr (NST) acceptor site for *N*-linked glycosylation is moved from a position 63 residues to a position 73 residues away from the P-site. The degree of glycosylation is measured for each translation product. *N*-glycosylation of a nascent chain is detected by an increase in molecular mass of about 2.5 kDa relative to the observed molecular mass of the non-modified molecule.





**Figure 24 | Cartoon representation of the approach.**

The model protein used in this study (*E. coli* Lep) has two TMD segments (gray) and a large C-terminal domain. Ribosome-bound truncated nascent chains of different lengths are generated by *in vitro* translation, in the presence of dog pancreas microsomes, of mRNAs (brown) lacking a stop codon. The minimum number of residues required to span the distance between the ribosomal P-site and the active site of the OST ( $d$ , distance P-NST) will depend on the compactness of the polypeptide region located inside the ribosome tunnel.

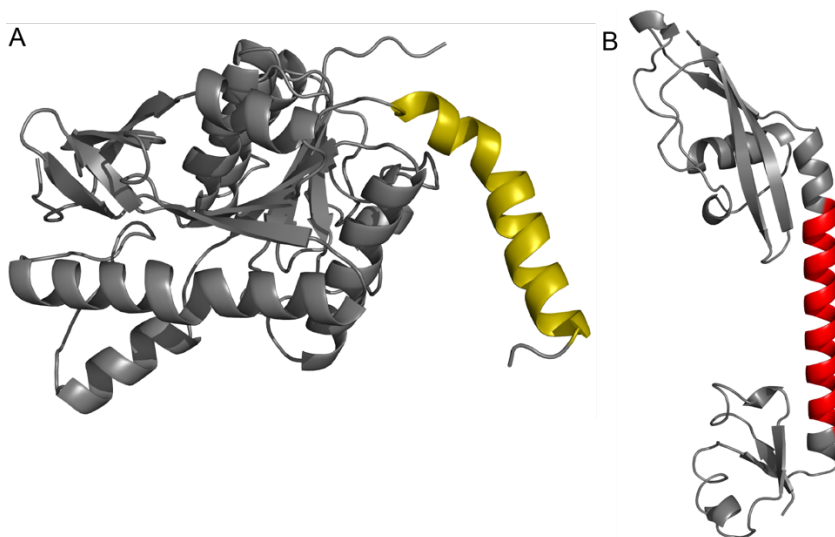
### c.1.2. TMD but not soluble helices have a compact conformation

It has been previously shown that a minimum distance of  $\sim 64$  residues from the C-terminus of a tethered nascent chain is required to bridge the P-site and the OST active site for sequences with extended conformation (from the extramembranous C-terminal domain of wild-type Lep). This P-NST distance is increased to  $\sim 70$  residues when model hydrophobic helical stretches are analyzed, suggesting a more compact, likely  $\alpha$ -helical conformation (Mingarro et al. 2000; Whitley et al. 1996). For a fully extended nascent chain ( $\sim 3.4$  Å per residue), more than 12 residues need to be folded to an  $\alpha$ -helix ( $\sim 1.5$  Å per residue) to account for this observed change in P-NST distance of  $\sim 6$  residues. Compaction of the nascent chain positions the glycosylation acceptor site closer to the membrane so that the acceptor Asp is no longer accessible to the OST active site (Figure 24). This arrangement is in line with the recently reported structure of mammalian ribosome-Sec61-OST complexes (Braunger et al. 2018).

In the current chapter, glycosylation mapping experiments were performed for nascent chains containing native helical sequences from the VSV-G protein or gp41 TMDs (Annex I). Glycosylation profiles were obtained that suggested compacted conformations (Figure 26A, upper panels). Nascent chains harboring non-TMD (soluble) helices of comparable lengths (Figure 25), either from an exceptionally stable helix from ribosomal protein L9 (Kuhlman et al. 1997) or from a highly hydrophilic *N*-acetylglutamate kinase (NAGK) (Ramón-Maiques et al. 2006) (Figure 26A, bottom panels), however, displayed a glycosylation pattern suggestive of an

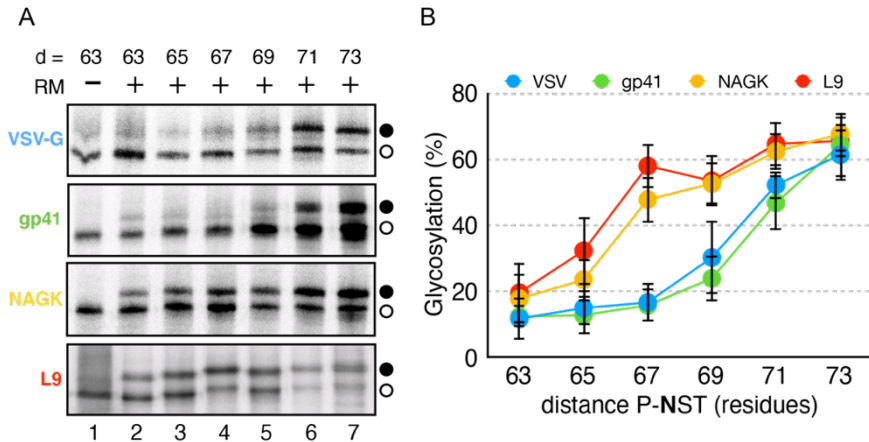
## Results

extended conformation for these sequences. These striking differences between the glycosylation patterns for the two types of helical sequences (Figure 26B) indicate that TMD helices may fold inside the ribosome exit tunnel, while nascent polypeptide chains harboring soluble helical sequences remain in an extended conformation. It should be noted that the folding event occurs far inside the ribosome exit tunnel (proximal to P-site), as the putative helix-forming sequences present in our constructs for P-NST distances of 67 residues are located at 7-9 residues from the C-terminus of the peptidyl-tRNA (Annex I). To demonstrate that the ribosomally non-compacting helical sequences from NAGK and L9 are soluble and not capable of inserting into the microsomal membranes in our experimental system, these constructs were analyzed using the Lep assay, a well-established assay to quantify the efficiency of membrane integration of tested sequences (Hessa et al. 2005). As expected, translation products of both these constructs revealed no membrane insertion (Figure 27).



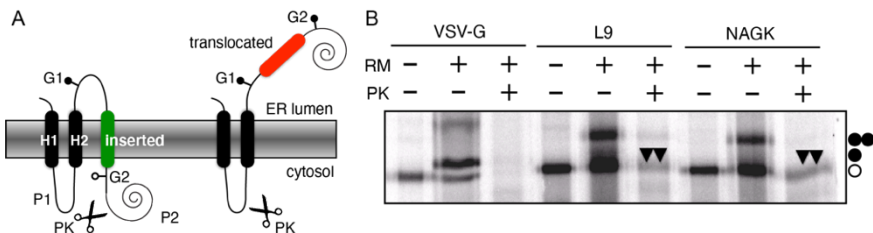
**Figure 25 | Atomic structure of the soluble proteins studied.**

(A) Cartoon representation of N-Acetylglutamate Kinase (NAGK) protein structure (PDB code 2BUF). The helical residues studied in this work are highlighted in yellow (from R5 to V26). (B) Cartoon representation of Ribosomal L9 protein structure (PCB code 1DIV). The helical residues studied are highlighted in red (K45 to K67).



**Figure 26 | Glycosylation profile of assayed sequences.**

(A) *In vitro* translation in the absence (-) or presence (+) of rough dog pancreas microsomes (RM) of truncated mRNAs of different lengths harboring the sequences encoding different helices: VSV-G TMD (residues 463-482), gp41 TMD (residues 684-705), NAGK helix (residues 5-26), and L9 helix (residues 45-67). The number of residues between the Asn residue in the Asn-Ser-Thr glycosylation acceptor site and the C-terminal end of the nascent chain is shown on top. Glycosylated and non-glycosylated molecules are indicated by black and white dots, respectively. (B) Glycosylation profiles for constructs of the indicated lengths harboring the different helical sequences. Error bars represent the mean  $\pm$  SD; obtained from at least 3 independent replicates.

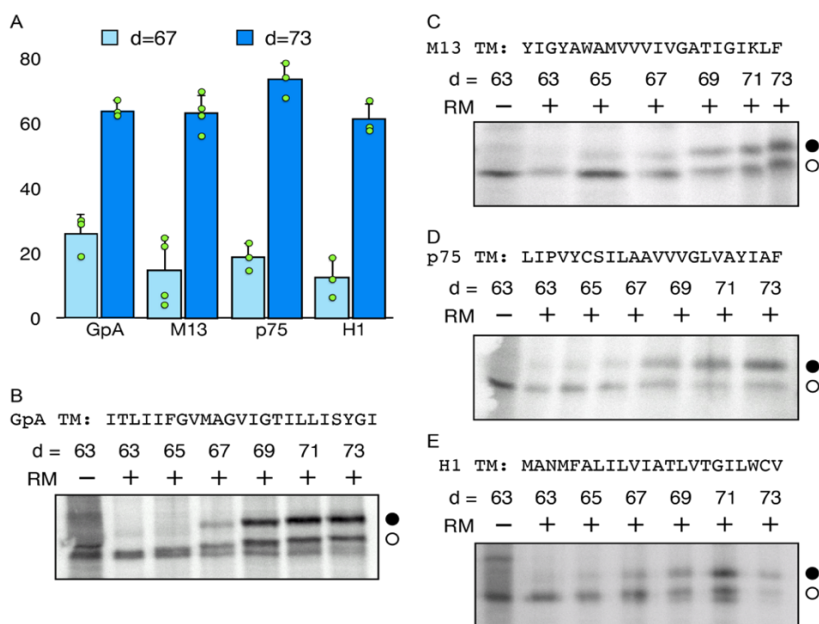


**Figure 27 | Insertion of hydrophobic regions of VSV-G, L9, and NAGK into membranes using Lep as model protein.**

(A) Schematic representation of the Lep construct used to report insertion of the hydrophobic region of VSV, and helical sequences of L9 and NAGK into ER membranes. The TMD under investigation (colored) was introduced into the P2 domain of Lep, flanked by two artificial glycosylation acceptor sites (G1 and G2). Recognition of the tested sequences as a TMD by the translocon machinery results in the location of only G1 in the luminal side of the ER membrane, preventing G2 glycosylation (left). The lep chimera will be doubly glycosylated when the sequence being tested is translocated into the lumen of the microsomes (right). (B) *In vitro* translation in the presence of membranes of the different Lep constructs. Constructs containing VSV-TM, L9 helix, and NAGK helix were translated in the presence (+) and absence (-) of rough microsomes (RM) and proteinase K (PK). Bands of non-glycosylated proteins are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. The protected doubly-glycosylated H2/L9/P2 fragments are indicated by two arrowheads.

## Results

To test whether the folding into helices in the ribosome is a common feature for TMDs, the study was extended to include the sequences of glycoprotein A TMD (GpA), the first TMD from Lep (H1), the TMD of p75 neurotrophin receptor (p75) and the TMD of the small coat protein of M13 phage (M13). As the maximal differential effect on glycosylation was previously observed for truncated Lep constructs occurred at a P-NST distance ( $d$ ) of 67 residues, this distance was defined as the critical number of amino acids required to distinguish between extended and compact conformations. These values were compared with a P-NST distance of 73, which is long enough to be fully glycosylated in all the constructs (Mingarro et al. 2000). All four constructs harboring TMD sequences show a clear difference in the glycosylation efficiency for both  $d=67$  and  $d=73$  distances (Figure 28A), being consistent with the adoption of an  $\alpha$ -helical structure in the ribosome tunnel of these tested sequences. In fact, the observed glycosylation patterns (Figure 28B-E) mirror those obtained for the VSV-G and gp41 constructs (Figure 26A, upper panels).



**Figure 28 | Glycosylation profile of TMD helices from different origins.**

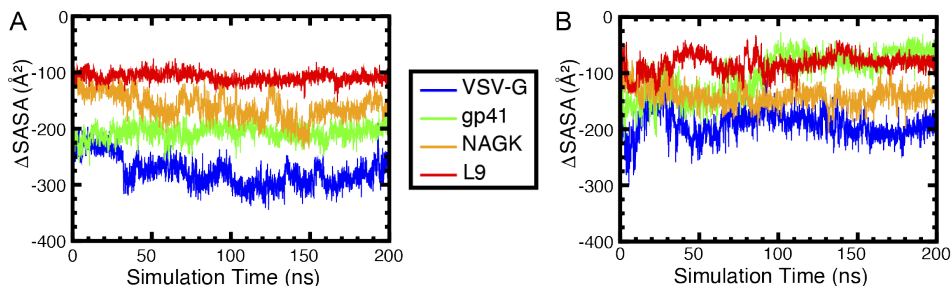
(A) TMD helices from different origins display a compact conformation. Glycosylation percentage of nascent polypeptides with  $d=67$  (light blue) or  $d=73$  (blue) residues between the acceptor Asn and the polypeptide C-terminus. Error bars represent the mean  $\pm$  SD;  $n \geq 3$ . Individual data points are shown as green dots. (B-E). Sequences and full glycosylation patterns of the assayed sequences. The TMD sequences used in each construct is shown on top of the respective SDS-PAGE autoradiography. Constructs harboring GpA (A), M13 coat protein (B), neurotrophin receptor p75 (C), and Lep H1 (D) were translated *in vitro* in the absence (-) and presence (+) of rough dog pancreas microsomes (RM). The 3' codon in the truncated mRNA was placed  $d$  codons (distances 63 to 73) downstream of the Asn residue in the Asn-Ser-Thr glycosylation acceptor site. Nonglycosylated and glycosylated molecules are indicated by white and black dots, respectively.

### c.1.3. Molecular dynamics simulation helix folding inside the mammalian ribosome

To gain more detailed insight into how mammalian ribosome exit tunnel distinguish between both helical and extended states, atomistic MD simulations were performed by the Gumbarts's Lab from the Georgia Institute of Technology (<https://simbac.gatech.edu>). Two hundred ns simulations of the 67-residue nascent peptide sequences VSV-G, gp41, NAGK, and L9 inside the mammalian ribosome exit tunnel starting from both helical and extended states were carried out. Although the conformational space of the peptide is unlikely to have been fully sampled in 200 ns, the initial position of the nascent polypeptide was based on that already present in the cryo-EM structure (Behrmann et al. 2015), which is similar to that of a nascent peptide in a translating ribosome (Gumbart et al. 2012). For all systems,  $\Delta$ SASA was measured. SASA is the reduction in the solvent-accessible surface area of the hydrophobic residues within the region of the nascent polypeptide known to be  $\alpha$ -helical in its final folded form due to contacts with the hydrophobic residues in the ribosome exit tunnel. Thus,  $\Delta$ SASA represents the degree to which hydrophobic contacts within the tunnels stabilize the nascent polypeptide compared to water; these contacts lower its free energy by  $\sim 0.015$  kcal/mol  $\cdot$   $\text{\AA}^2$  (Vallone et al. 1998). A more negative  $\Delta$ SASA indicates that the nascent polypeptide is more stable within the exit tunnel.

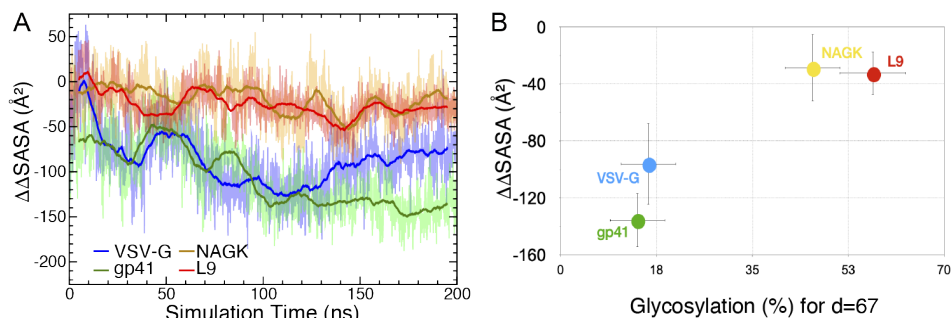
The values of  $\Delta$ SASA for both helical and extended states for all four nascent polypeptides inside the ribosome exit tunnel are plotted in Figure 29. To compare the relative stability of the helical state for the four nascent peptide sequences, the difference between  $\Delta$ SASA in the helical and extended states was calculated:  $\Delta\Delta$ SASA =  $\Delta$ SASA<sub>helical</sub> -  $\Delta$ SASA<sub>extended</sub>. After 100 ns,  $\Delta\Delta$ SASA for the TMD and soluble sequences begin to separate (Figure 30A). For the TMD sequences gp41 and VSV-G,  $\Delta\Delta$ SASA is negative, indicating that the helical state is stabilized by hydrophobic contacts in the ribosome exit tunnel more than the extended state. For the soluble sequences NAGK and L9,  $\Delta\Delta$ SASA is close to zero or only slightly negative, indicating that the extended state is neither favored nor disfavored compared to the folded state. To facilitate the interpretation of the folding results the glycosylation values at d=67 were compared with the average  $\Delta\Delta$ SASA (Figure 30B). Two different populations can be observed in the plot, corresponding to those which can acquire a compact conformation inside the ribosome exit tunnel and those which cannot fold inside the exit tunnel.

## Results



**Figure 29 | Solvent accessible surface area (SASA) for TMD and soluble sequences in the ribosome exit tunnel.**

$\Delta$ SASA for hydrophobic residues within nascent peptide sequences VSV-G (blue), gp41 (green), AGK (orange), and L9 (red). More negative values indicate more hydrophobic contacts with the ribosome. (A)  $\Delta$ SASA for  $\alpha$ -helical conformations. (B)  $\Delta$ SASA for extended conformations.



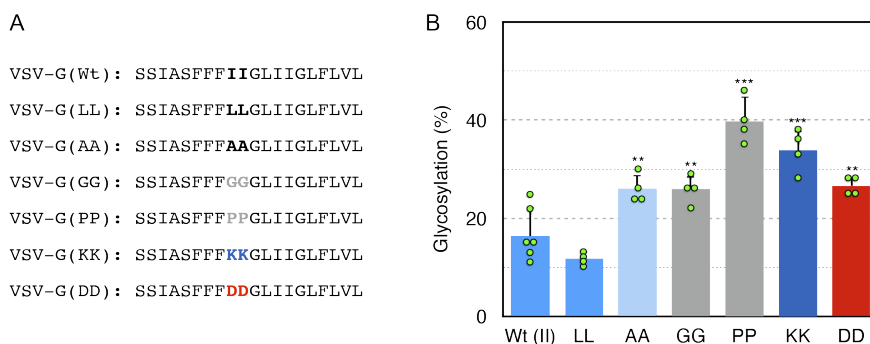
**Figure 30 | Solvent accessible surface area (SASA) for folded versus extended states.**

(A) Effect of ribosome solvent accessible surface area (SASA) of hydrophobic residues within the  $\alpha$ -helical sequences for folded versus extended states.  $\Delta\Delta$ SASA =  $\Delta$ SASA<sub>folded</sub> -  $\Delta$ SASA<sub>extended</sub>. Negative  $\Delta\Delta$ SASA values indicate that the ribosome is stabilizing hydrophobic regions of the helical sequence in the folded,  $\alpha$ -helical (compact) state more than in the extended state. Conversely, positive values indicate that the ribosome is stabilizing hydrophobic regions of the helical sequence in the extended state more than in the  $\alpha$ -helical state. (B) Scatter plot of the  $\Delta\Delta$ SASA versus the glycosylation profile for P-NST distance of 67. High percentage of glycosylation means no compact conformation, whereas a low percentage of glycosylation translates as compact conformation.

### c.1.4. Determinants of TMD helix folding inside the ribosome

To further study the determinants of TMD folding, the mid-region of the VSV-G TMD sequence was mutated to affect either hydrophobicity or helicity. When the central Ile pair (Figure 31A) was conservatively replaced with a Leu pair, no significant difference in glycosylation efficiency at P-NST distance of 67 residues

was observed (Figure 31B). However, reducing the hydrophobicity by mutating to an Ala pair resulted in a significant increase in glycosylation efficiency, indicating a less compact conformation. The increase in glycosylation efficiency was even more pronounced when charged residues were engineered into the sequence. Hence, constructs containing either Lys or Asp pairs displayed a glycosylation level indicative of a more extended conformation (Figure 31). Similarly, when the central Ile pair was replaced by helix breaking residues, either a Gly pair or a Pro pair, an increased level of glycosylation was observed, leading the greatest increase to the Pro pair. Based on results, it can be concluded that TMD sequence folding inside the ribosome exit tunnel depends on both hydrophobicity and helicity.



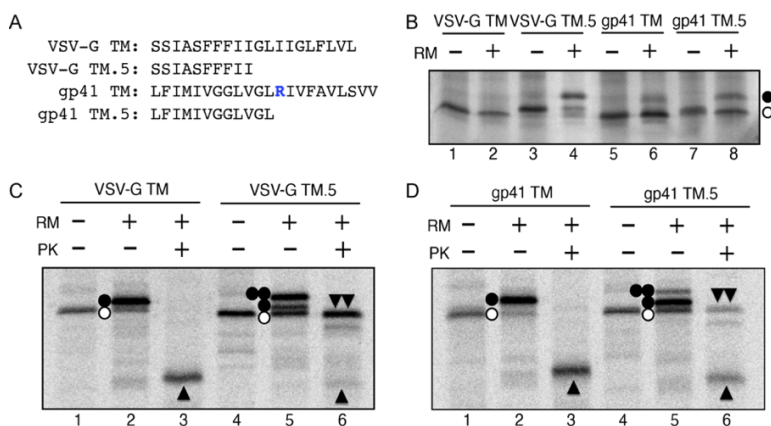
**Figure 31 | Hydrophobicity and helicity affect TMD folding.**

*In vitro* translation of truncated VS-G constructs (distance P-NST of 67) in which the central Ile pair (bold) was mutated to less hydrophobic, charged (basic and acid residues shown in dark blue and red, respectively) and helix breaking residues (shown in gray) (A). (B) Average glycosylation percentage for each mutant. Error bars show the standard deviation of four or more independent experiments (p-value for the comparison with wild type Ile pair: \*\* > 0.01 and \*\*\* > 0.001). Individual data points are shown as green dots.

If TMD sequences are responsible for the folding in the ribosome detected by glycosylation mapping, it is possible that the length of the hydrophobic sequence may influence folding events. The contribution to the overall length of a canonical  $\alpha$ -helix per amino acid is 1.5 Å. therefore, a stretch of ~20 consecutive hydrophobic amino acid residues are required to span the 30 Å of the hydrocarbon core of a ‘typical’ biological membrane (Grau et al. 2017). Indeed, the most prevalent length for TMD helices is 21 amino acids, according to structure-based statistical analysis (Baeza-Delgado et al. 2013). To investigate the relevance of the integrity of the TMD sequences in terms of length, constructs were designed containing half of the TMD sequence of VSV-G or gp41. As shown in Figure 32A, when truncated nascent chains of the same length (P-NST distance of 67 residues) containing only half of VSV-G TMD sequence (VSV-G TM.5) were translated in the presence of microsomal membranes efficient glycosylation was observed (lane 4). A similar observation was made for the construct harboring half of the hydrophobic sequence of the gp41 TMD,

## Results

(gp41 TM.5, compare lanes 6 and 8 in Figure 32B). These data suggest that short hydrophobic sequences adopt a more extended conformation. Hence, the ribosome exit tunnel can apparently distinguish between legitimate TMDs and a shorter stretch (~10) of non-polar residues in a nascent polypeptide. It should be noted, however, that the position of the short hydrophobic sequences in the ribosomal tunnel will likely be further away from the P-site than in the original TMD sequences due to the potentially extended conformation of the C-terminal residues.



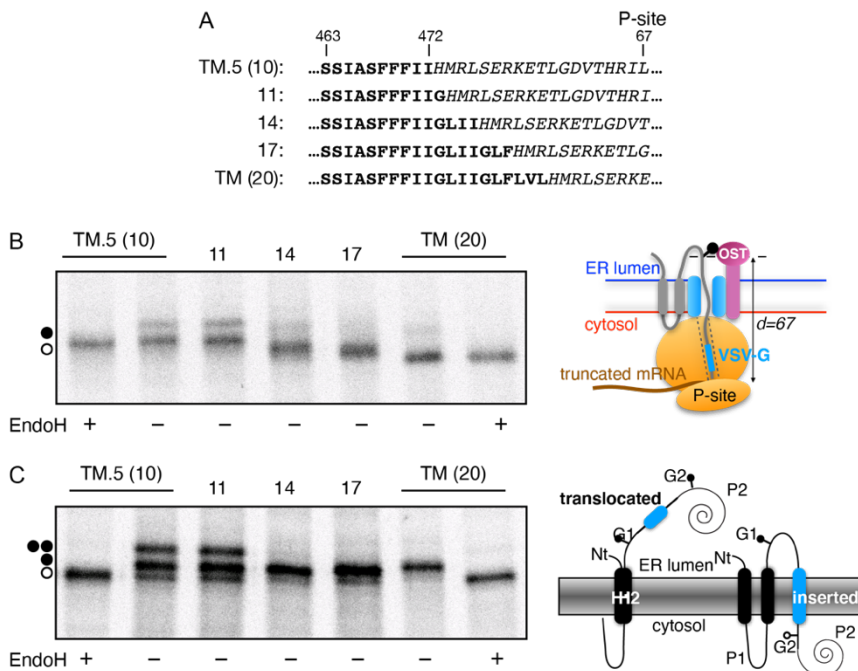
**Figure 32 | Folding depends on hydrophobic length and correlates with insertion.**

(A) TMD amino acid sequence cloned in the constructs. (B) *In vitro* translation in the absence (-) and presence (+) of dog pancreas rough microsomes (RM) of truncated mRNAs of the same length (distance P-NST of 67 residues) harboring VSV-G full length (lanes 1 and 2) or half TMD (VSV-G TM.5, lanes 3 and 4), or gp41 full length (lanes 5 and 6) or half TMD (gp41 TM.5, lanes 7 and 8). Glycosylated and non-glycosylated molecules are indicated by black and white dots, respectively. (C-D). *In vitro* translation of the Lep assay in the presence (+) or absence (-) of rough microsomes (RM) and Proteinase P (PK) of VSV-G (C) or gp41 (D) derived sequences. Non-glycosylated protein bands are indicated by a white dot; single and double glycosylated protein bands are indicated by one or two black dots, respectively. An upwards black triangle indicates a small protected singly glycosylated H2/inserted fragment. A double downward black triangle indicated a large doubly glycosylated H2/G1/translocated/G2/P2 fragment.

Finally, to investigate any correlation between TMD helix folding and insertion the efficiency of membrane integration of these sequences into microsomal membranes was measured using the Lep system (Hessa et al. 2005). In the presence of microsomal vesicles, the translation of constructs harboring the complete VSV-G TMD resulted in singly glycosylated forms of the protein (Figure 32C, lane 2). Proteinase K treatment of these samples yielded a small protected singly glycosylated H2/G1/inserted fragment (lane 3), indicating that the full-length VSV TMD sequence was properly inserted into the membrane. Comparable results were obtained with constructs harboring the gp41 complete TMD sequence (Figure 32D, lanes 1-3). However, the translation of constructs containing only half of VSV-G sequence



(VSV-G TM.5) in the presence of microsomes resulted mainly in double glycosylated (translocated) forms (69%, (Figure 32C, lane 5). Furthermore, following proteinase K treatment gave as result a largely protected fragment (H2/G1/translocated/G2/P2). With the gp41 sequence, it was found that 12 hydrophobic residues are sufficient for integration (Figure 32D, lanes 1-3). A correlation between TMD folding and TMD membrane insertion is also shown in Figure 33 where the VSV-G TMD acquires packing capabilities when the amino acid length is enough (at least 14 hydrophobic residues from the VSV-G TMD sequence) to promote the membrane insertion. These results suggest that a long hydrophobic stretch of amino acids is a requirement to facilitate stable helix formation in the ribosome. Thus, the folding of TMD helices appears to precede engagement with the translocon and subsequent insertion into the membrane.



(A) Subcloned VSV-G TMD sequences in the Lep system (highlighted in bold), TM.5 (residues 463-472), 11 (residues 463-473), 14 (463-476), 17 (463-479), TMD (full, residues 463-482) (B) *In vitro* translation in presence of rough microsomes and in presence (+) or absence (-) of Endoglycosidase H (Endo H, a glycan-removing enzyme) of the different truncated constructs. The 3' codon in the truncated mRNAs were placed 67 residues downstream of the Asn residue of the glycosylation acceptor site. Cartoon representation is shown on the right. (C) *In vitro* protein translation in presence of microsomes and in presence (+) or absence (-) of EndoH of the full sequence Lep containing the VSV-G TMD variations. Non-glycosylated protein bands are indicated by a white dot; single and double glycosylated protein bands are indicated by one or two black dots, respectively. Cartoon representation is shown on the right.

## Results

## Chapter 2. Interfacial scale determination

The characterization of an amphipathic or interfacial helix (IH) of a MPs can be a challenging task due to the difficulty to define common characteristics in terms of length and amino acid composition (see i.4.4). The analysis of interfacial helices is facilitated at some point by bioinformatic tools such as Heliquet (Gautier et al. 2008), a web server that screens sequences with specific  $\alpha$ -helical properties. This software is based on helical wheel projections and estimates parameters such as hydrophobicity, net charge, and hydrophobic moment. The hydrophobic moment is a useful index to quantify the amphiphilicity of a helix and to identify helical regions of proteins with amphipathic character (Eisenberg et al. 1982). Despite the bioinformatic efforts to identify and characterize IHs, they are still away to provide a global solution due to the absence of a specialized tool to detect IH formation and the aforementioned differences in composition between different IHs.

Experimental data acquisition about interfacial propensities for peptide sequences is not easy too. Structures of membrane-binding interfacial helices are scarce because of the difficulty in crystallizing proteins in an interfacial environment. Currently, a mixture of spectroscopic and biochemical methods can give reasonable clues about the relevance of a predicted interfacial helix. Some of the used techniques include liposome-protein binding assays, circular dichroism (CD) spectroscopy, or fluorescence methods. Despite this, one difficulty in the study of those helices is the finding of a model membrane system that is simple enough to be compatible with physical measurements, and yet similar enough to an actual membrane interface.

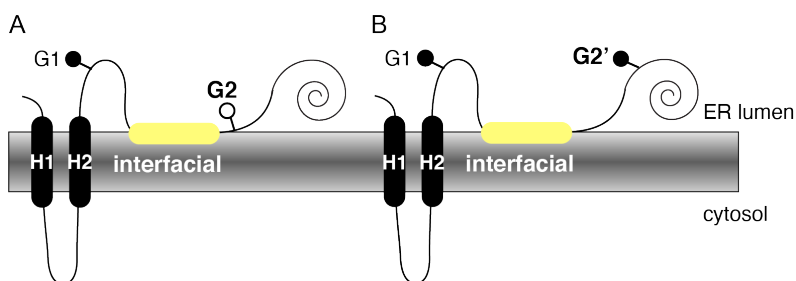
There is a wide variety of methods and algorithms to test whether a hydrophobic region will be inserted or not into the lipid bilayer, but an equivalent tool to rapidly measure the interfacial propensity of a particular sequence is not yet available. To gain the experimental information required to develop appropriate computational tools, a novel experimental design would be necessary. In the present thesis, a variation of the well-established 'Lep' model is proposed. In order to optimize the Lep assay to this end, glycosylation acceptor sites were meticulously engineered around the sequence under study, providing precise data about its final disposition into the membrane.

### c.2.1. Experimental design

The Lep-based assay (Hessa et al. 2005; Hessa et al. 2007) was modified by placing the second glycosylation (G2) acceptor site to different locations along the protein sequence (Figure 34). The basic idea behind this approach is that the OST can transfer

## Results

a glycosyl moiety to a nascent MP only when the acceptor site is placed at a minimum distance (number of residues away) from the membrane interface, the so-called minimal glycosylation distance (MGD) (Nilsson & G. von Heijne 1993). Then, the rationale behind our approach is as follows: if a sequence under study places interfacially, a sufficiently close G2 glycosylation acceptor site to the C-terminus of the interfacial segment will not reach the OST active site, and consequently, it will not be glycosylated. On the contrary, if the second glycosylation site (G2') is placed downstream, sufficiently away from the interfacial segment it will receive the glycan moiety. It should be mentioned that if the tested sequence is not retained at the interface but gets translocated, then both constructs the one harboring the glycosylation site close to the tested region (Figure 34, left) and the one carrying the G2' acceptor site (Figure 34, right), would be doubly glycosylated.



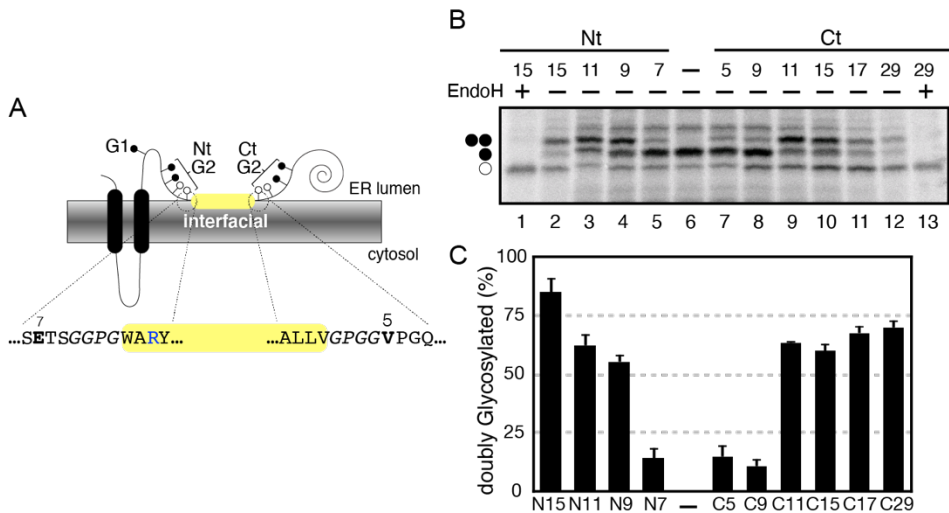
**Figure 34 | Schematic representation of the Lep variants.**

(A) Lep construct for the analysis of interfacial sequences with the G2 glycosylation site adjacent to the sequence under study (yellow rectangle). (B) Construct with the second glycosylation site (named as G2' to prevent confusion) away from the sequence under study.

To validate this assay, first the MGD for interfacial sequences should be determined. To carry out these studies, pHLIP3D, an interfacial sequence derived from the helix C of bacteriorhodopsin (Musial-Siwiek et al. 2010; Bañó-Polo et al. 2020) was used. The pHLIP3D peptide included an additional Asp residue between two inherent Asp residues ('3D') and the insertion of two Leu residues in the helix C sequence (Musial-Siwiek et al. 2010). Interestingly, CD studies in the presence of liposomes of the pHLIP3D peptide showed a surface helical configuration (Musial-Siwiek et al. 2010). Then, we have studied interfacial segments with the general design, GPGG-pHLIP3D-GPGG, in which the flanking tetrapeptides are included to insulate the central 24-residues (pHLIP3D) stretch from the surrounding sequence (Hessa et al. 2005) (sequences available in Annex II).

In our experimental setup, the G1 glycosylation site remains immovable as membrane reporter, while the second acceptor site (G2) was engineered at different positions in both N- or C-termini to the pHLIP3D sequence (Figure 35A). Translation of this set of constructs in the presence of microsomes displayed a glycosylation pattern revealing a threshold distance for N-linked glycosylation of 8-9 amino acid

residues for those constructs containing the G2 site preceding the interfacial sequence, and 10-11 amino acid residues for those constructs harboring the G2 site at the C-terminus of the pHLIP3D sequence (Figure 35B and C). Compared to the MGD found for membrane-spanning segments (14-15 residues upstream and 12-13 residues downstream) (Nilsson & G. von Heijne 1993), pHLIP3D interfacial sequence requires less distance from the two endpoints, likely because the sequence is placed at the luminal membrane interface, which means closer to the OST catalytic site.



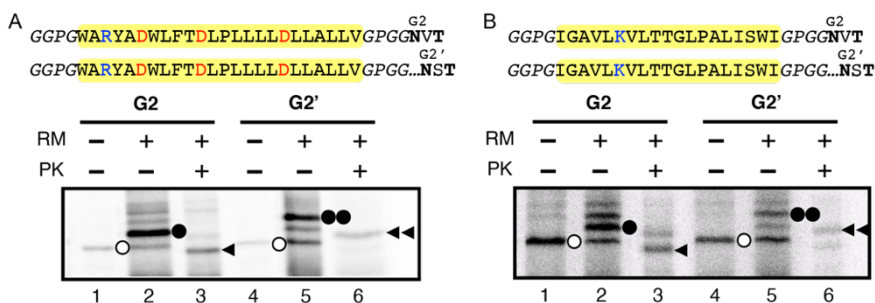
**Figure 35 | Determination of the N- and C-termini minimal glycosylation distance for interfacial sequences.**

(A) Schematic representation of the G2 glycosylation site location. The pHLIP3D sequence is represented with a yellow box. (B) *In vitro* translation of constructs with the G2 glycosylation site at different amino acid distances from both ends of the pHLIP3D sequence. Singly and doubly glycosylated forms of the protein are indicated with one or two black dots, respectively. The construct containing only G1 site (-, lane 6) was used as a control. EndoH + refers to the Endoglycosidase H treatment to determine the non-glycosylated form of the protein (white dot). (C) Glycosylation profile for constructs with the indicated distance between the interfacial sequence and the G2 glycosylation acceptor site (see Annex II for sequence details). Error bars represent the mean  $\pm$  SD; obtained from at least 3 independent replicates.

To corroborate the interfacial location of the pHLIP3D sequence in our experimental setup, appropriate glycosylation acceptor sites were placed in the construct. The original G1 position (Hessa et al. 2005) was kept at the original location, being our reporter for proper membrane targeting and insertion since it is placed more than 130 residues upstream of the interfacial sequence. The G2 acceptor site was placed to challenge the interfacial disposition of pHLIP3D sequence. For the constructs carrying G2 site (Figure 34), an NVT acceptor sequence was placed at position +5 downstream the interfacial sequence end thereby not been available for

## Results

glycosylation if the sequence locates at the membrane interface, but being only accessible to the OST if the tested sequence fully translocates to the microsomal lumen. On the contrary, to allow the glycosylation of the second acceptor site and to confirm that the interfacial sequence is not being inserted into the membrane, G2' containing constructs were fitted with the glycosylation acceptor site (NST sequence) at position +29 downstream the interfacial sequence ((Figure 35B, lanes 12 and 13, sequences available in Annex II). As proof of concept, the interfacial sequences of pHLIP3D (Figure 36A) and melittin (Figure 36B) were tested in both Lep-derived variants, G2 and G2'. As expected, the glycosylation profile obtained for the pHLIP3D sequence suggested an interfacial disposition (Figure 36A). When translated in the presence of rough microsomes (RM) the G2 construct profile was mainly singly glycosylated (lane 2), whereas G2' construct showed a doubly glycosylated pattern (lane 5). A proteinase K assay was performed to corroborate the results, being the polypeptide partially protected to the protease activity in both, G2 and G2' variants (lanes 3 and 6). Melittin sequence, a widely recognized interfacial peptide (Hristova et al. 2001), was also assayed in the modified Lep system, obtaining glycosylation patterns similar to those observed for pHLIP3D sequence (Figure 36B).



**Figure 36 | Interfacial Lep assay for pHLIP3D and melittin sequences.**

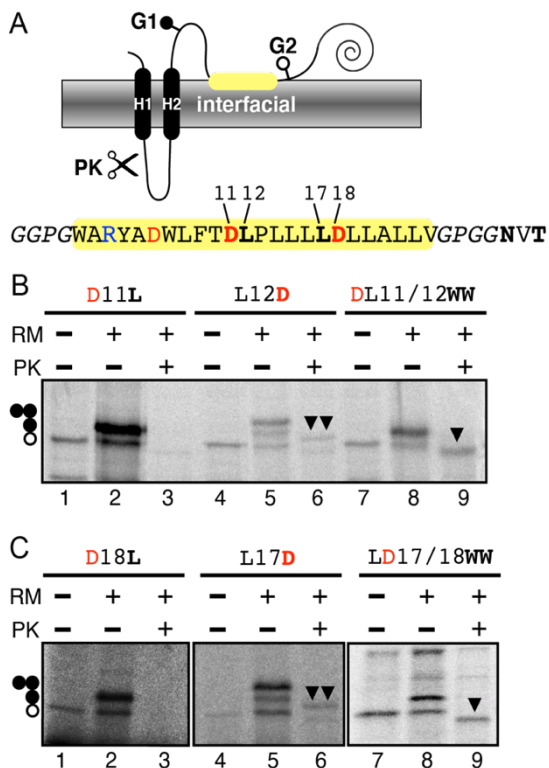
*In vitro* translation in the absence (-) or presence (+) of rough dog pancreas microsomes (RM) of G2 and G2' constructs for pHLIP3D (A) and melittin (B). Interfacial sequences are highlighted in yellow. Bands of non-glycosylated proteins are indicated by a white dot; singly and doubly glycosylated proteins are indicated by one and two black dots, respectively. The protected singly- or doubly-glycosylated H2/G1/pHLIP3D/P2 or H2/G1Melittin/P2 fragments from proteinase K assay (PK) are indicated by one or two arrowheads, respectively.

### c.2.2. pHLIP3D replacement points

We next sought to determine the naturally occurring amino acid contribution to the final interfacial propensity of a given sequence. To perform these experiments, the starting point was the screening of the pHLIP3D sequence in order to find an optimal location where amino acid substitutions could cause a global and measurable modification in terms of interfaciality of the whole sequence. Based on the initial pHLIP3D sequence, in which the authors replaced the original bacteriorhodopsin T90 with an Asp and inserted a Leu pair (Musial-Siwiek et al. 2010), we decided to screen those positions corresponding to D11 and L12 of the pHLIP3D sequence. To search for the best location for our systematic analysis the equivalent L17 and D18 residues were also included in the screening. To this end, amino acid replacements were performed in the constructs harboring the G2 glycosylation acceptor site (Figure 37A). Then, a pair of high hydrophobic residues (Leu-Leu), a pair of polar charged residues (Asp-Asp), and a pair of aromatic residues found highly interfacial on previous scales (Wimley & White 1996) (Trp-Trp) were designed (sequences available in Annex II). Those three residue pairs potentially provide all possible options of the sequence disposition in the membrane. When challenged the more centered positions (11 and 12) (Figure 37B), the LL pair (obtained by D11L mutation) showed singly-glycosylated molecules (lane 2) and no proteinase K protected band (lane 3), suggesting that the presence of two consecutive Leu residues in the positions 11/12 leads to efficient sequence insertion. For the DD pair (obtained by L12D mutation) a mostly double glycosylated profile (lane 5) was observed together with a proteinase K protected band (lane 6), suggesting sequence translocation. The WW pair also showed a singly-glycosylated protein band (lane 8) but, in contrast to the LL construct, proteinase K treatment rendered a protected band (lane 9), suggesting an interfacial disposition for this construct. When equivalent substitutions were engineered at 17 and 18 positions (Figure 37C), similar results were observed, confirming that the pHLIP3D sequence is able to provide a suitable vehicle to study the interfacial contribution of any single amino acid along its sequence, since point mutations successfully switch its final membrane disposition.

All in all, for the rest of the work included in this chapter we selected the more centered 11 and 12 position since the mutations on positions 17 and 18 pair could affect the helix folding due to its implications in the initial helix folding of the pHLIP peptide (Scott et al. 2017).

## Results



**Figure 37 | pHLIP3D sequence screening for systematic substitutions.**

(A) Schematic representation of the Lep G2 with the interfacial sequence pHLIP3D (highlighted in yellow). The pHLIP3D amino acid pairs tested are highlighted in bold (DL11/12 pair and LD17/18 pair). *In vitro* translation in the presence (+) or absence (-) of dog pancreas rough microsomes (RM) of the plasmids harboring the DL11/12 (B) or the LD17/18 (C) pair substitutions. Non-glycosylated molecules are indicated with a white dot while single and double glycosylated protein bands are indicated with one or two black dots, respectively. A downwards black triangle indicates large protected singly glycosylated H2/G1/interfacial/P2 fragment. A double downward black triangle indicated a large doubly glycosylated H2/G1/translocated/G2/P2 fragment. Negatively and positively charged residues are shown in red and blue, respectively.

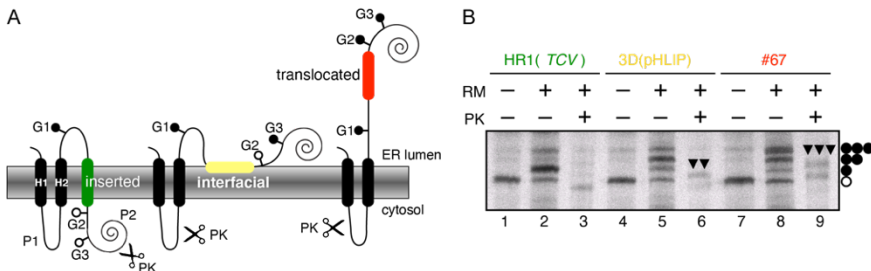
### c.2.3. Improvement of the methodology by inserting a third glycosylation acceptor site

For a best performing and to facilitate the study of putative interfacial sequences on a unique scaffold, we designed a variation of the glycosylation assay where the 3 different glycosylation acceptor sites (G1, G2, and G2') were included in the same construct (named as Lep3G). To facilitate data interpretation, the glycosylation acceptor sites were named in order of appearance in the sequence, which means that G1 and G2 acceptor sites were kept (Figure 37A) and the G2' acceptor site was renamed as G3 acceptor site (C29 position, Figure 35), placing all of them within the same construct (Figure 38A). With this design, the glycosylation pattern varies significantly, allowing data interpretation from a single construct. Hence, the single-glycosylated molecules keep as indicative of sequence insertion into the membrane (only the G1 acceptor site can be glycosylated), whereas the translocate state is characterized by the glycosylation of the 3 possible acceptor sites simultaneously



(G1, G2, and G3). Thus, the interfacial state could be differentiated by the presence of doubly-glycosylated molecules, in which only sites G1 and G3 will expect to be glycosylated (Figure 38A).

To validate our design, 3 different sequences were cloned into Lep3G plasmid and assayed. Those sequences correspond to: 1) HR1 TM, a highly hydrophobic sequence from Turnip Crinkle Virus (TCV) movement protein, previously demonstrated to insert efficiently through the translocon into microsomal membranes (Martínez-Gil et al. 2010); 2) pHLIP3D sequence as shown above with a neat interfacial behavior; 3), a pseudo-randomized and previously tested sequence with a high propensity to be translocated named #67 (Sääf et al. 1998; Peiró et al. 2014) (Figure 38B). As expected, the HR1 from TCV was mostly singly-glycosylated (lane 2) and digested after proteinase K treatment (lane 3), suggesting efficient membrane insertion. Interestingly, the interfacial pHLIP3D sequence showed a prevailing double glycosylation pattern (lane 5), with a proteinase K protected band corresponding to the region H2/G1/pHLIP3D/G3/P2 (lane 6). Predictably, the translocation control sequence (#67) was mostly triply glycosylated and rendered a proteinase K protected band after PK treatment that matches with the region H2/G1/translocated/G2/G3/P2 (lane 8 and 9, respectively).



**Figure 38 | Interfacial LepG3 assay.**

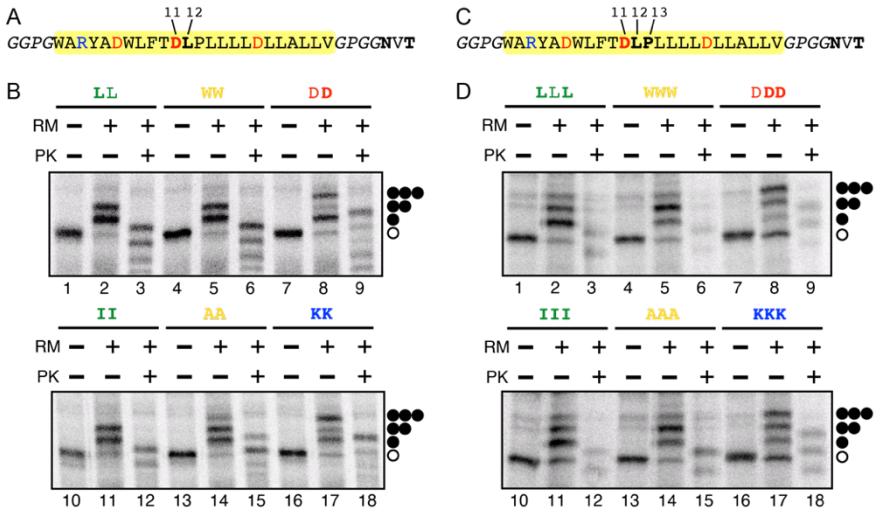
(A) Schematic representation of the LepG3 design where all three glycosylation acceptor sites (G1, G2, and G3) are placed in the same construct, displaying a glycosylation pattern reflecting the final disposition of the assayed sequence (inserted, interfacial, or translocated). (B) *In vitro* translation in the absence (-) or presence (+) of rough microsomes (RM) of LepG3 constructs harboring the first TMD for Turnip Crinkle Virus movement protein (HR1 TCV), interfacial sequence (pHLIP3D) and translocated pseudo-randomized sequence from (Sääf et al. 1998)(#67). Bands of non-glycosylated proteins are indicated by a white dot; single, double, and triple glycosylated proteins are indicated by one, two, or three black dots, respectively. The protected double or triple glycosylated H2/G1/pHLIP3D/G2/G3/P2 and H2/G1/#67/G2/G3/P2 fragments from proteinase K assay (PK) are indicated by two or three arrowheads, respectively.

Once demonstrated that pHLIP3D is a suitable scaffold to challenge sequence interfaciality using the Lep3G assay, naturally occurring amino acid substitutions were performed in the DL11/12 residues of pHLIP3D sequence. Firstly, 6 different pairs of amino acids were engineered replacing the original Asp-Leu pair at 11/12

## Results

positions (Figure 39A). The selected pairs were two hydrophobic residues (Leu and Ile), two residues with high interfacial propensity (Trp and Ala) according to previous studies (Wimley & White 1996; Nilsson et al. 2003), and two charged residues with opposite polarity (Asp and Lys) (Sequences available in Annex III). As in the sequence screening using the G2 and G2' Lep variants (Figure 37), these substitutions modified the final disposition of the chimeric proteins in the microsomal membranes. Constructs harboring hydrophobic residues Leu and Ile exhibited a mostly single glycosylation pattern (Figure 39B, lanes 2 and 11 respectively) indicative of membrane insertion. The interfacial residues Trp and Ala displayed glycosylation profiles quite similar to those obtained for the hydrophobic residues (Figure 39B, compare lanes 5 and 14 to lanes 2 and 11). On the contrary, charged residues showed a clear triple glycosylated pattern suggesting protein translocation, though a significant fraction of glycosylated molecules remained singly glycosylated (Figure 39B, lanes 8 and 17).

To magnify the differences between amino acid substitutions the replacement of 3 consecutive residues was considered. To this end, 11, 12, and 13 positions corresponding to Asp-Leu-Pro (DLP) tripeptide in the pHLIP3D sequence (Figure 39C) were replaced by triplets of the same aforementioned amino acids (Sequences available in Annex III). As expected, triplet substitutions showed a clearer glycosylation pattern. For the hydrophobic residues Leu and Ile, single glycosylated molecules were found prevalent (Figure 39D, lanes 2 and 11), whereas the interfacial residues Trp and Ala mainly displayed a double-glycosylated pattern (Figure 39D, lanes 5 and 14), strongly suggesting a surface disposition in the membrane. For constructs harboring charged residues, a triply glycosylated band was predominant in the glycosylation profile (Figure 39D, lanes 8 and 17), suggesting the translocation of these sequences. Based on these results, the triplet substitutions DLP 11/12/13 appeared to maximize differences between naturally occurring amino acid residues in the glycosylation profile compared to pair substitutions (Figure 39 panels B and D).



**Figure 39 | Amino acid substitutions in DL pair and DLP triplet.**

Interfacial sequence pHLIP3D (highlighted in yellow). The pHLIP3D amino acid pairs subjected to replacement are marked in bold (DL11/12 pair (A) and DLP11/12/13 triplet (B)). *In vitro* translation in the absence (-) or presence (+) of rough dog pancreas microsomes (RM) of pHLIP3D substituted amino acids in DL11/12 pair (C) and DLP 11/12/13 triplet (D). Bands of non-glycosylated proteins are indicated by a white dot; single, double, and triple glycosylated proteins are indicated by one, two, or three black dots, respectively. Amino acid color code represents expected insertion (green), interfacial (yellow), or translocated (red or blue for negative or positive charged amino acids, respectively).

#### c.2.4. Calculating interfacial propensity for individual amino acids

Once the best target for amino acid substitution was selected, the degree of insertion, interfaciality, or translocation for any given sequence could be calculated from SDS-PAGE bands quantification by measuring the fraction of singly ( $f_{1g}$ ), doubly ( $f_{2g}$ ), and triply ( $f_{3g}$ ) glycosylated Lep-derived molecules. The next paragraphs are going to be focused on the equations involving the interfacial state, being the equivalents for the inserted and translocated states attached as Annex IV and Annex V, respectively.

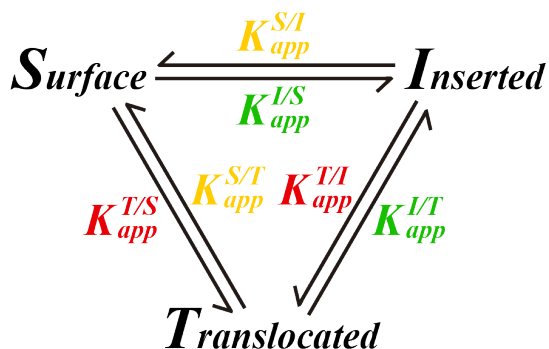
The interfacial (surface) propensity can be calculated using the Lep3G assay as the probability of the sequence to be double glycosylated,  $p_{surface} = f_{2g}/(f_{1g} + f_{2g} + f_{3g})$ . These data can also be expressed as an apparent equilibrium constant ( $K_{app}$ ) between the surface disposition ( $S$ ) and the summed alternative states (i.e.,  $I$  for membrane inserted and  $T$  for membrane translocated) (Figure 40). Since 3 states are possible, 2 different  $K_{app}$  should be calculated for the surface disposition.

## Results

One involving the inserted state ( $K_{app}^{S/I} = K_{app}^{I/S^{-1}} = f_{2g}/f_{1g}$ ) and the other involving the translocated state ( $K_{app}^{S/T} = K_{app}^{T/S^{-1}} = f_{2g}/f_{3g}$ ) dispositions. Note that for any given apparent equilibrium constant  $K_{app}^{X/Y}$  there exists a reverse expression  $K_{app}^{Y/X}$  ( $K_{app}^{X/Y} = K_{app}^{Y/X^{-1}}$ ) that represents the opposite equilibrium between both states (Figure 40). As previously described (Hessa et al. 2005), those results can be easily converted to apparent free energies for direct comparison with biophysical data:

$\Delta G_{app}^{S/I} = -\Delta G_{app}^{I/S} = -RT \ln K_{app}^{S/I}$ ;  $\Delta G_{app}^{S/T} = -\Delta G_{app}^{T/S} = -RT \ln K_{app}^{S/T}$ , where  $R$  is the gas constant (in kcal/mol) and  $T$  is the thermodynamic temperature of the reaction (in Kelvin units)

For  $\Delta G_{app}$  values, 0 represents the perfect equilibrium between the two states involved in each case, whereas non-zero values mean preference for one of the two states in the equilibrium. Given the equation  $\Delta G_{app}^{X/Y} = -RT \ln K_{app}^{X/Y}$ , negative values will represent a preference for state X, while state Y preference will be represented by positive values.



**Figure 40 | Schematic representation of the equilibrium among the three possible states.**

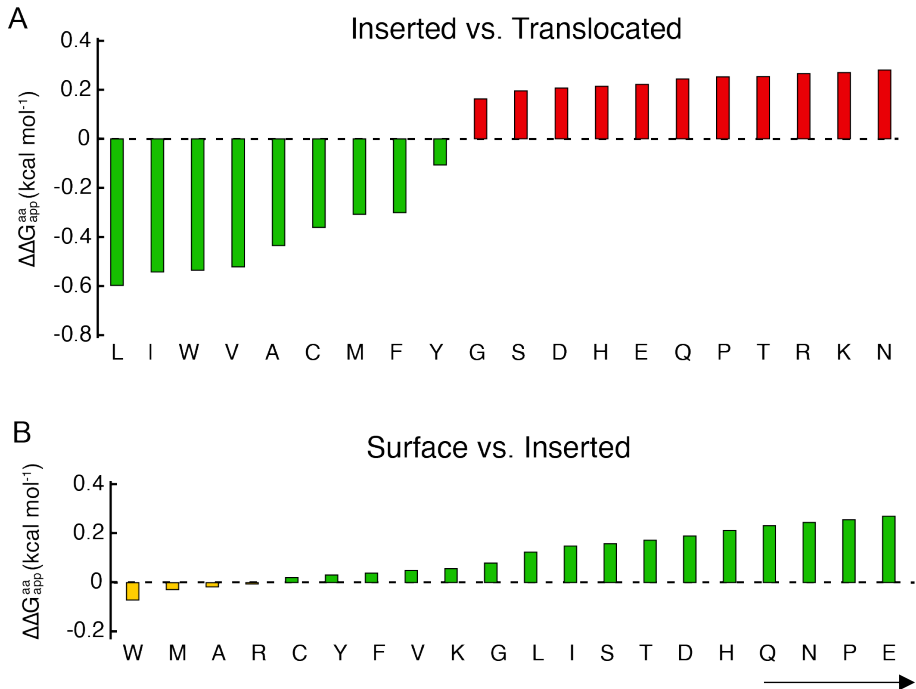
Green, yellow and, red color code was used to highlight the equilibrium constant that drives to inserted, surface, and translocated state, respectively.

To determine the amino acid contribution, the difference of free energy of the sequence with DLP11/12/13X substituted amino acid ( $\Delta G_{app}^{xxx}$ ) could be calculated as mentioned above. Once the  $\Delta G_{app}^{xxx}$  is calculated, the difference for individual amino acid contribution with respect to the DLP triplet ( $\Delta \Delta G_{app}^{x-DLP}$ ) can be derived from the difference of the free energy of the pHLIP3D ( $\Delta G_{app}^{WT}$ ) with respect to the free energy of the analyzed sequence with substituted amino acids:

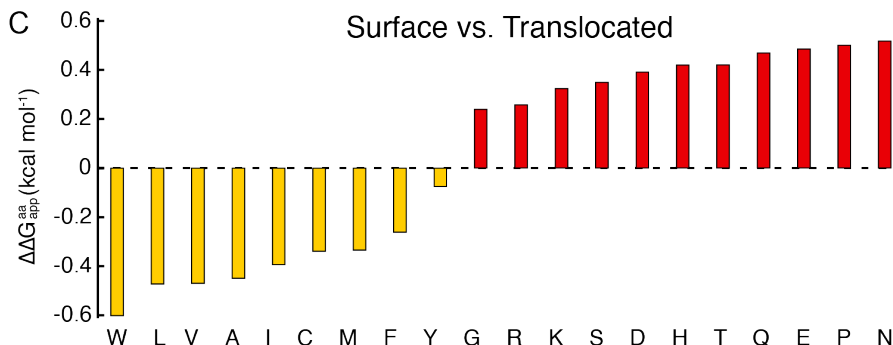
$$3\Delta \Delta G_{app}^{x-DLP} = \Delta \Delta G_{app}^{xxx-WT} = \Delta G_{app}^{xxx} - \Delta G_{app}^{WT};$$

$$\Delta G_{app}^{x-DLP} = \frac{\Delta G_{app}^{xxx} - \Delta G_{app}^{WT}}{3}$$

The data obtained for  $\Delta\Delta G_{app}^{aa}$  (also known as  $\Delta\Delta G_{app}^{x-DLP}$  along the calculations) of insertion, surface, and translocation states for the DLP replacement with the 20 naturally occurring amino acids (sequences available in Annex III) was represented in 3 different scales (Figure 41). The values were calculated from individual constant apparent equilibria:  $K_{app}^{I/T}$  for the inserted/translocated equilibrium (Figure 41A),  $K_{app}^{S/I}$  for surface/inserted (Figure 41B), and  $K_{app}^{S/T}$  for surface/translocated (Figure 41C). The plots displayed the behavior of any single amino acid when analyzing the equilibrium between two of the three putatively different states. For example, negative  $\Delta\Delta G_{app}^{Ile}$  for the inserted/translocated equilibrium means a preference for membrane insertion against translocation. As opposite, positive  $\Delta\Delta G_{app}^{Lys}$  for the same equilibrium reflects Lys membrane translocation preference when compared to membrane insertion.



## Results



**Figure 41 | Apparent free energy amino acid scale from individual equilibrium constants.**

Since 3 different states are possible, 3 different equilibrium constants and apparent free energies could be calculated for any given amino acid: Inserted vs translocated state,  $K_{app}^{I/T}$  (A), surface vs inserted state,  $K_{app}^{S/I}$  (B), and surface vs translocated state,  $K_{app}^{S/T}$  (C). Green, yellow and red color code was used to highlight amino acid propensity for inserted, surface or translocated state on each equilibrium, respectively

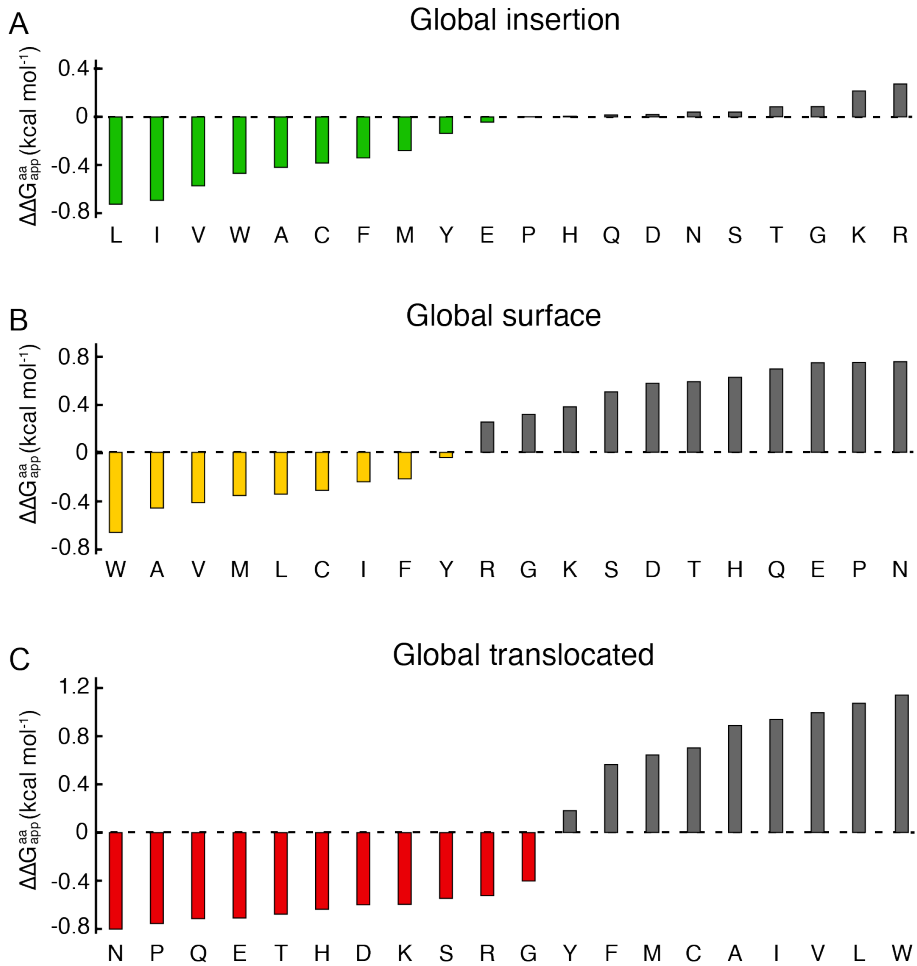
To get a comprehensive analysis of the amino acid's preferences, a global  $K_{app}$  can also be calculated as a merge of the two previous equilibrium constants, since each state is influenced by the other two (Figure 40):  $K_{app}^{surface} = K_{app}^{S/I} \cdot K_{app}^{S/T} = f_{2g}/f_{1g} \cdot f_{2g}/f_{3g}$ .

As previously,  $\Delta\Delta G_{app}$  can also be calculated by using the equation:  

$$\Delta\Delta G_{app}^{surface} = \Delta\Delta G_{app}^{S/I} + \Delta\Delta G_{app}^{S/T} = -RT \ln K_{app}^{surface}$$

The  $\Delta\Delta G_{app}^{aa}$  values obtained correspond to the merge of the two apparent free energies in which each state is implied (e.g., the global surface apparent free energy for a Trp residue will be:  $\Delta\Delta G_{app}^{surface Trp} = \Delta\Delta G_{app}^{S/I Trp} + \Delta\Delta G_{app}^{S/T Trp}$ ). These  $\Delta\Delta G_{app}^{aa}$  values (Figure 42) simplify the visualization of the individual amino acid global tendency. Negative global  $\Delta\Delta G_{app}^{aa}$  values for a given state means a propensity to this state for each residue type. On the contrary, positive values indicate a displacement of the equilibrium to the other two possible states. The analysis of those three global residue  $\Delta\Delta G_{app}^{aa}$  values (global insertion, global surface, and global translocated) will give us an accurate perspective for each amino acid behavior. For instance, negative  $\Delta\Delta G_{app}^{Ile}$  values can be observed in both scenarios, global inserted and global surface (Figure 42), the higher absolute value for the inserted state emphasizes isoleucine's preference for the inserted state over the surface location. To accurately understand this behavior, an analysis of  $\Delta\Delta G_{app}^{Ile}$  value for surface vs. inserted equilibrium (Figure 41B) is needed. The observed Ile values emphasize its higher propensity for the inserted state. By contrast, the soaring positive global

translocated  $\Delta\Delta G_{app}^{Ile}$  value indicates the reluctance of Ile residues to be translocated across the membrane.

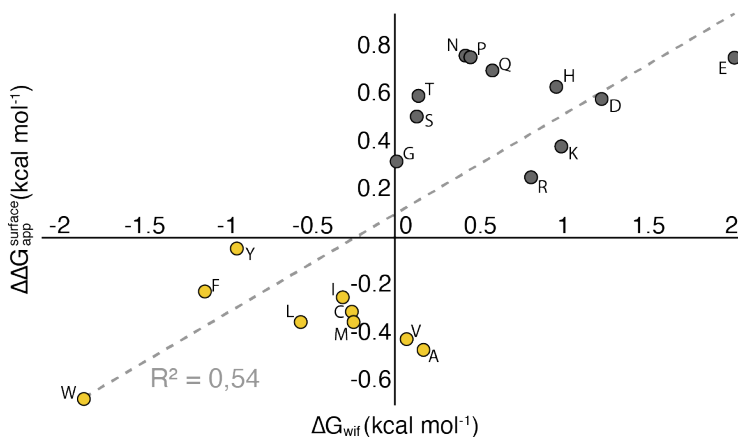


**Figure 42 | Amino acid apparent free energy values for each state.**

$\Delta\Delta G_{app}^{aa}$  values correspond to the merge of the two apparent free energies in which each state is implied for any given combination (global insertion = inserted/surface + inserted/translocated; global surface = surface/inserted + surface/translocated; global translocated = translocated/inserted + translocated/surface). Green, yellow, and red color code was used to highlight amino acid propensity for insertion, interfaciality, or translocation on each equilibrium, respectively. Gray color bars correspond to positive  $\Delta G$  values, pinpointing a disfavored state in each case.

## Results

In any case, the current ‘biological’  $\Delta\Delta G_{app}^{surface}$  scale has to be compared with previous scales derived from biophysical measurements. The correlation between the  $\Delta\Delta G_{app}^{surface}$  scale and the Wimley-White water/POPC interface scale ( $\Delta G_{wif}$ ) (Wimley & White 1996) is shown in Figure 43.



**Figure 43 | Correlation between biological and biophysical scales.**

(A) Correlation between the ‘biological’  $\Delta\Delta G_{app}^{surface}$  scale (Figure 42B) and the biophysical Wimley-White water/POPC interface hydrophobicity scale ( $\Delta G_{wif}$ ) extracted from (Wimley & White 1996). Color code as in Figure 42B where yellow dots correspond to amino acid propensity for interfaciality. Gray color dots correspond to positive  $\Delta\Delta G_{app}^{surface}$  values, pinpointing a disfavored surface state.

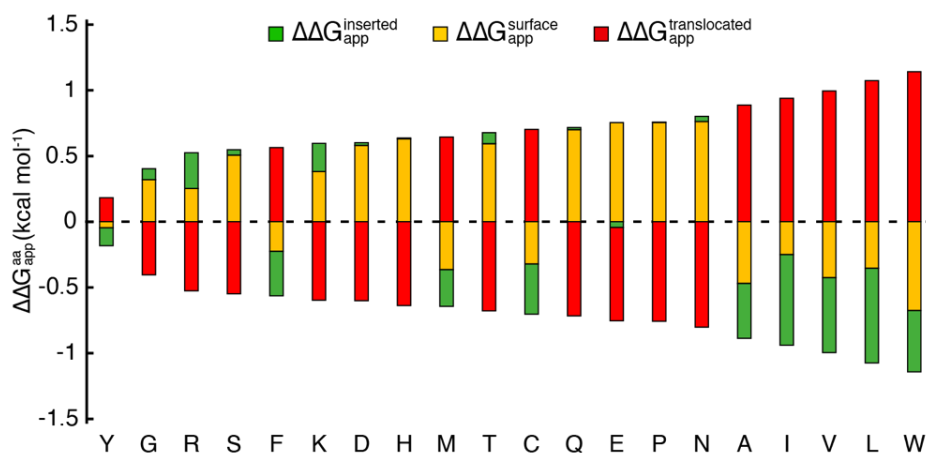
Since the three states’ equilibria for any given amino acid is a closed system, the sum of the three different global states  $\Delta\Delta G_{app}^{aa}$  should be zero, displaying a perfect relation between them. Based on this assumption, the combination of those 3 sets of values (one for each state) in a single chart will provide a more user-friendly visualization of the behavior of any single amino acid residue in the three states as shown in Figure 44. As before, when compared, positive  $\Delta\Delta G_{app}^{aa}$  values for any given state are indicative of a thermodynamic penalty for the particular amino acid toward that state. Conversely, negative  $\Delta\Delta G_{app}^{aa}$  values are indicative of spontaneous propensity toward those states. It should be noted that when negative  $\Delta\Delta G_{app}^{aa}$  values for two different states are found in the same amino acid, the propensity of the amino acid is towards the state with the higher absolute negative value.

As an example, the high positive  $\Delta\Delta G_{app}^{Trp}$  value for the translocated state means that Trp residues display a strong preference toward the inserted or surface states. Between the negative  $\Delta\Delta G_{app}^{Trp}$  values, both insertion and surface dispositions are found spontaneous. The final propensity of the amino acid will be towards the more



negative value, in the case of tryptophan, the surface state (Figure 44). This means that Trp residues will prefer a surface state according to the values found for the equilibrium between surface and inserted states (as a consequence of higher negative value, see also Figure 41B). It also has to be considered that, as an equilibrium, a certain tendency to the inserted state should be considered but being less prevalent than the propensity to adopt an interfacial location.

In terms absolute values, higher bars mean larger differences in propensity between different states, whereas shorter bars mean lower differences in propensity between different states. Tyr is the residue with lower absolute values, which means not a strong preference for any state. In contrast, Trp is the residue with higher absolute values, meaning that the residue has strong preferences for being inserted or interfacial and thus preventing the translocated state.



**Figure 44 | Global view for residue's apparent free energy.**

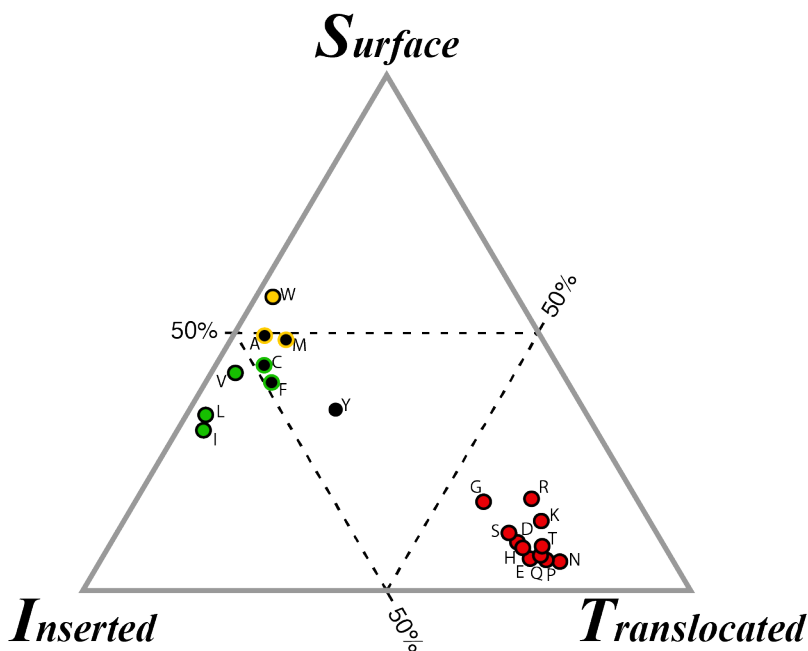
Green, yellow and red color code was used to highlight amino acid propensity for insertion, interfaciality or translocation on each amino acid, respectively. Negative values mean propensity for the given state whereas positive values are indicative of low propensity to the respective state. Amino acids are ordered based on the absolute free energy values. Higher bars are indicative of large differences in propensity among different states, whereas shorter bars are indicative of small differences.

Finally, amino acid propensity was plotted as the molar ratio for each state. To do so,  $K_{app}^{aa}$  should be calculated from  $\Delta\Delta G_{app}^{inserted}$ ,  $\Delta\Delta G_{app}^{surface}$ , and  $\Delta\Delta G_{app}^{translocated}$ .

$$K_{app}^{inserted} = e^{\frac{\Delta\Delta G_{app}^{inserted}}{-RT}}; K_{app}^{surface} = e^{\frac{\Delta\Delta G_{app}^{surface}}{-RT}}; K_{app}^{translocated} = e^{\frac{\Delta\Delta G_{app}^{translocated}}{-RT}}$$

## Results

Once the equilibrium constants are calculated, the molar ratio distribution could be estimated, since the final propensity of the amino acids is determined by these equilibriums. Figure 45 summarizes these calculations (probability to find a particular amino acid in any state) for any given amino acid in a triangular plot. As can be observed, Trp residue is embedded in the upper left side of the chart, which is characterized by a surface state preference. In the same way Ile, Leu, and Val amino acid residues are placed in the bottom left side of the chart, indicative of a preference for the membrane inserted state. Ala, Met, Cys, Phe, and Tyr residues are placed centrally, being characterized by a non-clear preference for any state. Within the amino acids included in this category, it was possible to find residues (Ala and Met) with certain tendency to be found at the interface and others (Cys and Phe) with some tendency by the inserted state, despite not showing a strong preference for them. Finally, the rest of the naturally occurring amino acid residues are placed in the bottom right vertex of the triangle, an area that includes charged and polar residues, which are characterized by a preference for the translocated state.



**Figure 45 | Individual amino acid propensity across the three different states.**

Triangle representation of the naturally occurring 20 amino acids, based on its states' preference. Amino acids found in the corner triangles are characterized by a major propensity to the labeled state (inserted (green dots), surface (yellow dot), or translocated (red dots) state). Amino acids placed in the central triangle showed not clear state preference (black dots). Black dots with green and yellow lines refer to residues with a certain tendency to the inserted and surface state, respectively.

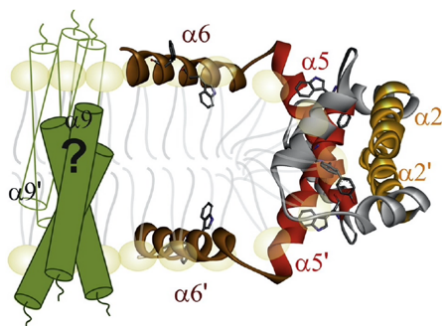
### c.2.5. Characterization of Bax-derived amphipathic peptide

Apoptosis is a programmed cell death process that plays a key role in the homeostasis of multicellular organisms. The pathways involved in the apoptosis process are complex, as well as the external signals that trigger the process (UV radiation, hormonal signals, or chemotherapy drugs, among others). Apoptosis is fundamentally regulated by the interplay between the BCL-2 (*B cell lymphoma-2*) family proteins. This BCL-2 family includes several pro- and anti-apoptotic proteins that trigger or prevent the apoptotic process, respectively. From a structural view, the BCL-2 family proteins conserve homology domains called BH, corresponding to several segments in  $\alpha$ -helix conformation (Chao & Korsmeyer 1998). The BH1-BH2 domain, where the  $\alpha$ 5/6 helices are found, has strong biological membrane affinity. In fact, it has been described that both Bax protein helices together (Bax  $\alpha$ 5/6) are capable to permeabilize membranes as a consequence of the membrane curvature they induce (Garcia-Saez et al. 2006). This phenomenon is also called *interfacial activity* and it is defined as the molecule's ability to attach the membrane in the water-lipid interface and disturb the lipid packing and organization (Fjell et al. 2007; Rathinakumar et al. 2009). The molecular mechanism by which peptide segments permeabilize membranes remains unclear, but it is known that non-conserved sequence motifs are found in these peptides. As a general rule, the common characteristics in terms of their cationic composition and hydrophobicity (high net positive charge and amphipaticity) give them their membrane permeabilizing capacity (Wimley 2010). In the cell, the target of the BCL-2 family proteins is the mitochondrial outer membrane, where pro-apoptotic proteins permeabilize the membrane allowing the intermembrane content release and thus the caspase pathway activation, the true executor of the cell death (Basañez et al. 2012; Westphal et al. 2014). Besides, evidence has been found that this type of helices can permeabilize other types of membranes and release their contents (Kuwana et al. 2002; Epand et al. 2002; Andreu-Fernández et al. 2014).

Guided therapy development against several diseases is rising in recent years. These therapies take advantage of molecular particularities of the affected cells. In cancer cells, local acidosis is caused by metabolic alterations that lead the cell to produce glycolysis-derived acids as a consequence of the anaerobic metabolism (also known as Pasteur effect) (Sennoune et al. 2004). Recently this phenomenon has been exploited to develop an effective tool based on pH-sensitive peptides (pHLIP) to detect, tag, and deliver drugs to the surface of cancer cells (Andreev et al. 2009; Andreev et al. 2010; Wijesinghe et al. 2011). The design is based on the acidic characteristics of the aspartic acid and glutamic acid naturally present in the sequence (see a detailed explanation of pHLIP3D on page 71 and sequence in Annex III, which allow the peptide to adopt an  $\alpha$ -helix structure and inserts into the membrane in acid environments, being a promising property for

## Results

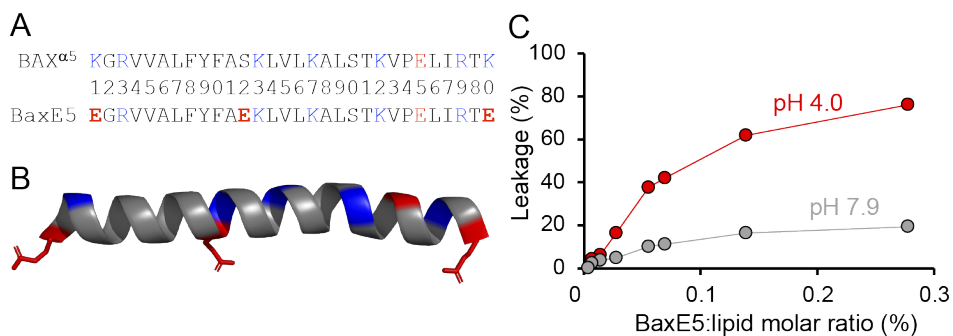
therapeutic applications, as diseased states characterized by extracellular acidosis include aggressive solid tumors (Estrella et al. 2013), arthritic inflammation (Andreev et al. 2007), and sepsis (Henry et al. 2020). Based on the described evidence and the membrane interfacial topology of the helices  $\alpha 5$  and  $\alpha 6$  in the active Bax (Figure 46), we thought about the possibility to modify a previously used pore-forming Bax <sup>$\alpha 5$</sup>  peptide (Garcia-Saez et al. 2006) (Figure 47A, top).



**Figure 46 | Model of active dimeric bax at the membrane**

Hypothetical clamp model for the topology of active Bax dimers at the Mitochondrial Outer Membrane (MOM). The dimerization domain is at the rim of a pore induced by Bax in the membrane, with helices  $\alpha 5$  and  $\alpha 6$  lying on the membrane surface. Taken from (Bleicken et al. 2014).

The peptide sequence was modified strategically introducing three glutamic acid residues at non-conserved positions (Xiao et al. 2016) (Figure 47A, bottom and B) to confer pH-response properties. In addition, Glu residues promote membrane translocation according to data shown above (Figure 42). The resulting BaxE5 peptide (Figure 47A) was readily soluble in aqueous solution at close to neutral pH. Since the parental Bax <sup>$\alpha 5$</sup>  peptide permeabilizes lipid bilayers in a concentration-dependent manner, the BaxE5 concentration effect on membrane integrity was measured by Sulforhodamine B leakage of large unilamellar vesicles (LUVs). As expected, BaxE5 can efficiently release Sulforhodamine B at low peptide concentrations (Figure 47C). Interestingly, the degree of dye leakage was controlled by pH. At neutral pH inefficient leakage was observed at the BaxE5 concentrations assayed. In contrast, BaxE5 efficiently disrupted membrane integrity at low concentration and acidic pH, in agreement with the design principle pursued.

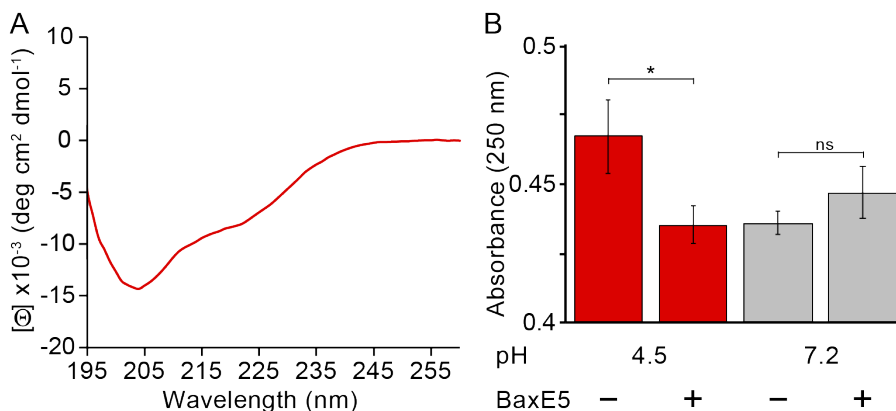


**Figure 47 | BaxE5 selectively disrupts lipid integrity.**

(A) Sequence comparison between Bax<sup>Δ5</sup> peptide and the modified BaxE5 peptide. Three acidic residues (Glu) were introduced into the sequence of the Bax<sup>Δ5</sup> at positions 1, 12, and 30 (highlighted in bold). Negatively charged Glu residues are shown in red and positively charged Lys and Arg are shown in blue. (B). Representation of a helical BaxE5 peptide structure. Blue colored structure represents positively charged Lys and Arg whereas red colored represents negatively charged Glu residues. The three designed Glu residues are represented with their side chains (sticks view). Image assembled in Pymol using the residues 106-125 of the structure of Bax (PDB code 4BD7), represented as cartoon and sticks. (C) Leakage of Sulforhodamine B encapsulated in POPC vesicles measured at pH 4.0 (red) and pH7.9 (gray) at different peptide concentrations, expressed as peptide:lipid molar ratio.

CD spectroscopy was used to determine the conformation of BaxE5 in presence of POPC lipids at acidic pH. The CD spectra had a minimum wavelength at 204 nm, a value typically observed for peptides and proteins in a largely unstructured conformation (Kelly et al. 2005; Kelly & Price 2000) (Figure 48A). A shoulder was observed at 222 nm, which indicates that the peptide has a weak  $\alpha$ -helical character, in contrast to the original peptide that in aqueous solvent displayed values close to 40% total helical structure, considering both regular and distorted  $\alpha$ -helix (Garcia-Saez et al. 2006). Furthermore, at pH 4.5, the presence of BaxE5 reduced the absorbance at 250 nm of the sample (POPC LUVs), but no effect was observed at neutral pH (Figure 48B, gray bars). These changes in absorbance are possibly the result of a reduction of the light scattering in the presence of LUVs and suggest that BaxE5 at acidic pH induces a disruption in the vesicle structure.

## Results



**Figure 48 | BaxE5 disrupts membranes in a largely unstructured conformation.**

(A) Circular dichroism spectrum of BaxE5 peptide in presence of POPC vesicles at pH 4.3. The molar lipid to peptide ratio was 200 to 1. (B) Effect the presence (+) or absence (-) of BaxE5 peptide in the absorbance at 250 nm of LUVs at pH 4.5 (red colored bars) or 7.2 (gray colored bars). Error bars represent the mean  $\pm$  SD; obtained from at least 3 independent replicates. p-values for the comparison with peptide and non-peptide treatment: \*  $<0.05$ , ns non-significant.

According to the results obtained, we can conclude that the rationally modified BaxE5 peptide disturbs the integrity of lipid membranes, being the effect strongly dependent on pH. A severe membrane disruption only takes place at acidic pH at nanomolar concentrations of the peptide (ranging from 3 to 500 nM). The absorbance reading at 250 nm resulted in changes in vesicle light scattering dependent on the presence of peptide and acidic pH conditions. Since scatter is proportional to vesicle size, the absorbance reduction observed at acidic pH suggests that BaxE5 disrupts the integrity of lipid vesicles.

### Chapter 3. The role of hydrophobic matching on TMD helix packing

Assembly of the native structure of most integral MPs takes place in two main steps (Popot & Engelman 1990). The first step includes targeting and insertion of the protein into a lipid membrane. In the case of  $\alpha$ -helical MPs this initial step occurs generally co-translationally through the translocon. In the second state, if required, the TMD interact to form the tertiary and quaternary structure of the mature functional membrane protein.

While for water-soluble proteins the dynamics and energetics underlying the folding have been studied thoroughly, the extent of similar studies in the context of MPs is much more modest. This is quite surprising, given the abundance of MPs in the human proteome and its important roles. Due to their different environments, the forces that underlie the folding process are also distinct for water- and membrane-soluble proteins (Martínez-Gil et al. 2011; Hong 2014; Martínez-Gil & Mingarro 2015). For water-soluble proteins, the folding is largely driven by hydrophobic interactions. In the folding of MPs, the role of the hydrophobic effect is less relevant and applies mainly to the formation of secondary structures (Cymer et al. 2015). Also, while salt bridges and aromatic interactions are important in the folding of water-soluble proteins, they do not contribute significantly to the MP folding (Bañó-Polo et al. 2012). Meanwhile, there are forces such as inter-helical hydrogen bonding and especially van der Waals interactions that have only a minor role in the folding of soluble proteins, while there are considered a major driving forces in protein folding within lipid bilayers (Martínez-Gil et al. 2011; Martínez-Gil & Mingarro 2015).

One of the means used by membranes to control TMD conformation is hydrophobic matching, i.e., the matching between the hydrophobic span of a TMD and the hydrophobic thickness of the lipid membrane around the proteins (Mouritsen & Bloom 1984; Andersen & Koeppe 2007). Given that exposure of hydrophobic groups in proteins and lipids to water is highly unfavorable, membranes tend to minimize their free energy by maximizing the matching between the hydrophobic length of the bilayer and the TMD helices. However, in some cases, there is a disparity thus creating a hydrophobic mismatch. The resulting energetic penalty is thought to be compensated either by membrane or peptide rearrangements, including TMD packing (Killian 1998). Intriguingly, while this concept has been explored quite extensively for individual TMD (peptides) in model membranes and also under *in vitro* conditions (Jensen & Mouritsen 2004; Lee 2005), it has received much less attention in the more realistic setting of living cells.

## Results

GpA represents one of the best-suited and most studied models for  $\alpha$ -helical TMD packing and membrane protein folding (MacKenzie 2006; Orzáez et al. 2000). GpA homo-dimerization relies exclusively on its unique TMD (Lemmon et al. 1992). The sequence motif within the TMD driving the association can be reduced to five residues, namely G<sub>79</sub>VxxGVxxT<sub>87</sub> (where x represents any hydrophobic residue) (Orzáez et al. 2005). Amidst this motif, the glycine residues play a crucial role. Their disposition, coupled with the tilt of the helix, renders close packing of two monomers, thereby maximizing significant interactions between the TMD (DeGrado et al. 2003). However, experimental results have shown that, at least *in vitro*, the formation of GpA dimers is not solely dependent on the protein sequence. The lipid environment can also make a significant contribution (Kuznetsov et al. 2015). It has been shown that not only the lipid composition but also the hydrophobic mismatch between the GpA TMD and the surrounding hydrophobic environment of the lipid membrane can modify the monomer-dimer equilibrium (Orzáez et al. 2005). The above view based on experimental work is supported by molecular simulations of model systems, where GpA has served as a centerpiece. Hence, MD simulations on GpA have been employed to investigate phenomena such as membrane insertion (Bond & Sansom 2006), dimer structure (Petrache et al. 2000; Flinner et al. 2014), and dimerization energetics (Sengupta & Marrink 2010).

Since its introduction in the nineties (Bloom et al. 1991), the concept of hydrophobic mismatch has received extensive attention both experimentally (*in vitro*) (Anbazhagan & Schneider 2010; Grau-Campistany et al. 2016; Muhle-Goll et al. 2012; Soubias et al. 2015) and computationally (de Jesus & T. W. Allen 2013; T. Kim & Im 2010; Kandasamy & Larson 2006). However, as mentioned above, the implications of hydrophobic matching on membrane protein folding, packing, and oligomerization have not been investigated in biological membranes of cells. To address it, we explore hydrophobic matching and its effects through GpA dimerization in prokaryotic and eukaryotic cells. To this end, fluorescence-based assays together with different length TMDs harboring the GpA dimerization motif were used for a better understanding of membrane protein folding under native conditions.

### c.3.1. Packing of TMD with different length in biological membranes

To assess the influence of hydrophobic matching on the packing of TMD, one should vary the length of either the TMD or the model membrane system employed in the assay. Working in cells, the only modifiable variable is the length of the TMD used. For this purpose, a series of chimeric stretches bearing the minimal dimerization



domain found in GpA with an increasing number of Leu forming TMDs were constructed (Table 1). All the hydrophobic regions designed were identified as TMD by the  $\Delta G$  prediction server (Hessa et al. 2007). The designed TMDs range from 17 to 29 residues long. The rise per residue along the axis in a canonical helix is 1.5 Å. Therefore, a stretch of approximately 20 consecutive hydrophobic amino acids will span a 30 Å of the hydrocarbon core of a biological membrane. Indeed, the most prevalent length of TMD helices is 21 residue (Baeza-Delgado et al. 2013). By selecting TMDs that are either longer or shorter than 21 residues a discrepancy in the membrane matching can be induced, allowing to investigate the role of this imbalance in the TMD packing.

**Table 1 | Sequences, predicted  $\Delta G$  and hydrophobic length of transmembrane segments**

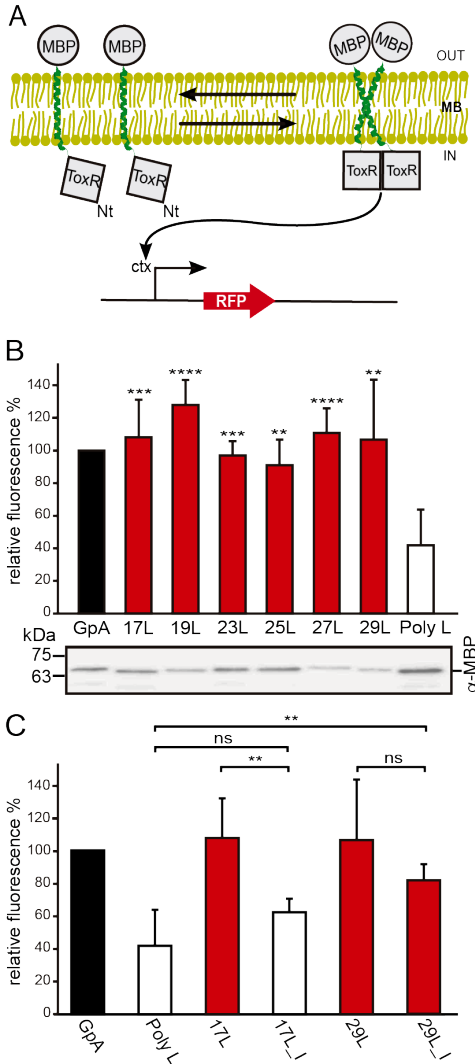
Chimeric TM	Sequence	$\Delta G_{app}^{pred}$ (kcal/mol)	Hydrophobic length (Å)
GpA	GRPNITLIIF <sup>79</sup> <b>GV</b> MAG <sup>83</sup> <b>VI</b> GTILLISYGIEYP	-1.646	34.5 Å
17L	GRPNLKLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLEYP	-1.688	25.5 Å
19L	GRPNLKLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLEYP	-3.469	28.5 Å
23L	GRPNLKLLLLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLLLEYP	-5.775	34.5 Å
25L	GRPNLKLLLLLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLLLEYP	-7.397	37.5 Å
27L	GRPNLKLLLLLLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLLLEYP	-8.698	40.5 Å
29L	GRPNLKLLLLLLLLLL <sup>79</sup> <b>GV</b> LL <sup>83</sup> <b>GV</b> LLTLLLLLEYP	-9.494	43.5 Å
17L_I	GRPNLKLLLL <sup>79</sup> <b>I</b> VLL <sup>83</sup> <b>I</b> VLLTLLLEYP	-3.847	25.5 Å
29L_I	GRPNLKLLLLLLLLLL <sup>79</sup> <b>I</b> VLL <sup>83</sup> <b>I</b> VLLTLLLLLEYP	-11.625	43.5 Å

Chimeric TMDs were named based on their hydrophobic length (amino acids). The sequence of each TMD is included. The dimerization motive is highlighted in bold (including the amino acid position in the wild-type GpA sequence) and flanking regions are indicated in gray. The apparent predicted  $\Delta G$  for the insertion of hydrophobic regions (calculated by the  $\Delta G$  prediction server v1.0, where negative values are indicative of insertion) and the hydrophobic length (calculated assuming 15 Å per residue in an  $\alpha$ -helix conformation) were also included in the table.

In order to understand the hydrophobic matching effect in a cellular environment, the ability of the aforementioned TMDs to homo-dimerize in *E. coli* membranes was firstly studied. To this end, a variation of the ToxCAT (Russ & Engelman 1999) assay known as ToxRED (Berger et al. 2010) was utilized. Briefly, this methodology uses a chimeric construct composed of the N-terminus DNA binding domain of ToxR (a dimerization-dependent transcriptional activator), fused to the challenged TMD and a periplasmic anchor (MBP) needed for the growth of the bacteria in minimal media supplemented with maltose (Andreu-Fernández et al.

## Results

2016). Dimerization through the TMDs results in ToxRED-mediated activation of the *ctx* promoter which drives the synthesis of the Red Fluorescent Protein (RFP) (Figure 49A). RFP values were normalized using the absorbance of the bacteria culture (600 nm) to rule out culture growth differences as the source of fluorescence variations (note that in this system the growth of the MM39 *E. coli* strain depends on the proper expression and insertion of the chimeric protein, due the endogenous MBP lack). Furthermore, the correct expression of all the constructs was assessed by western blot using an anti-MBP antibody (Figure 49B, bottom).

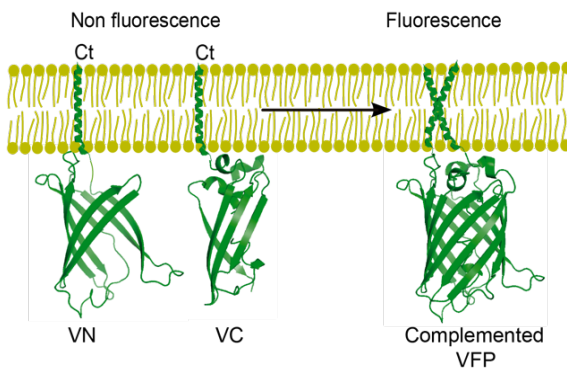


**Figure 49 | Homo-dimerization in *E. coli* membranes.**

(A) Schematic representation of ToxRED Assay. TM-driven oligomerization results in dimerization of ToxR transcriptional activator which, ultimately, drives the expression of the red fluorescent protein RFP (encoded under *ctx* promoter). The C-terminus maltose-binding protein (MBP) located in the periplasm (OUT) allows growth of *E. coli* MM39 strain in M9 minimal media supplemented with 0.8% of maltose. (B) Mean relative fluorescence of chimera homo-oligomerization. Error bars denote standard deviation obtained from at least 6 independent experiments (p-values for the comparison with poly L: \*\* <0.01, \*\*\* <0.001, \*\*\*\* <0.0001). The color intensity-code was used to highlight dimerization (red) vs non-dimerization (white). The positive control (GpA homo-dimer) is shown in black. The  $\alpha$ -MBP western blot under the bar graph shows chimera's expression levels. (C) Contribution of Gly for the dimerization of TMD chimeras. Mean relative fluorescence of 17L, 29L, 17L\_I, and 20L\_I chimeras homo-oligomerization. Error bars denote standard deviation obtained from at least 4 independent experiments (p-values for the comparison with poly L: \*\* <0.01, ns non-significant). The color intensity-code was used to highlight dimerization (red) vs non-dimerization (white). The positive control (GpA homo-dimer) is shown in black.

The results show that all the chimeras bearing the minimized GpA dimerization motif, despite their different hydrophobic lengths, were capable of forming homo-dimers that render RFP levels similar to those obtained using the wild-type GpA TMD (a 23 hydrophobic residue long segment), and significantly higher than the negative controls a 13 amino acid long stretch of Leu, poly L, that efficiently inserts into the membranes (Jaud et al. 2009; Baeza-Delgado et al. 2016) (Figure 49B). It has been shown that long hydrophobic segments can lead to oligomerization by themselves as a packing method (Li et al. 2004; Wei et al. 2011). To isolate the contribution of the hydrophobic length on the oligomerization, the TMDs constructs were mutated replacing the Gly residues to Ile in the 17L (17L\_I) and 29L (29L\_I) (Figure 49C). The sequence, hydrophobic length, and predicted  $\Delta G$  values of these segments are included in Table 1. While elimination of GxxxG motif in the 17L backbone decreased the ToxRED associated fluorescence to background levels, the Gly to Ile substitutions had a minor effect on 29L, indicating that the positive mismatch can induce TMD packing in *E. coli* membranes.

Additionally, the formation of homo-dimers in eukaryotic cells was analyzed utilizing a BiFC assay. Briefly, the Venus Fluorescent Protein (VFP) was divided into two non-fluorescent parts: amino-terminus (VN) and carboxyl-terminus (VC). Each half was then fused to the hydrophobic segments previously designed (Table 1) and expressed in human-derived HEK-293T cells as in (Andreu-Fernández et al. 2017). Oligomerization of the TMDs allows the reconstitution of the full-length VFP and the recovery of its fluorescence properties (Figure 50). Similarly to the ToxRED assay, native GpA TMD homo-dimers were included as a positive control and normalization value. As a negative control, the second TMD of *E. coli* Leader peptidase (H2), a non-dimerizing protein widely used in membrane protein biogenesis studies (White & G. von Heijne 2008), was used.

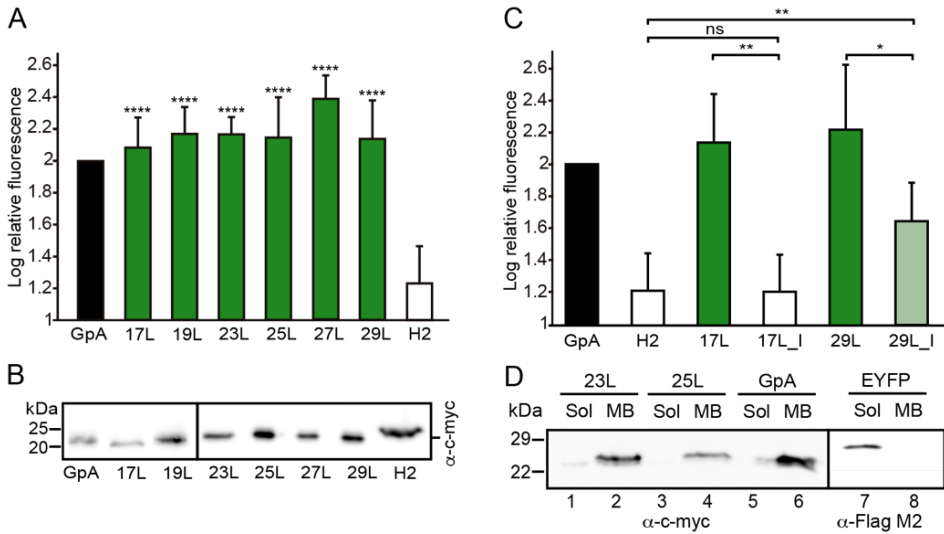


**Figure 50 | Schematic representation of BiFC Assay.** TM-driven oligomerization results in complementation of two non-fluorescent halves (amino-terminus (VN) and carboxy-terminus (VC)) of the Venus Fluorescent Protein (VFP).

## Results

In eukaryotic membranes, as in bacterial membranes, a strong dimerization of all tested GpA-based chimeras was observed. Regardless of their hydrophobic length (from 17L to 29L) the fluorescence values were comparable to those obtained when the wild-type TMD of GpA was used (Figure 51A). A western blot was included to monitor protein levels (Figure 51B). The contribution of the GpA dimerization domain to the interaction between TMD monomers was also investigated in the BiFC assay. Once again, Gly residues were substituted to Ile in the 17L and 29L constructs (VN and VC). In both cases, in contrast to the ToxRED results, the elimination of the Gly residues significantly reduced the observed fluorescence (Figure 51C). Nonetheless, a fluorescence increase of the 29L\_I could be observed as compared to the 17L\_I, suggesting that positive but not negative hydrophobic mismatch can partially drive oligomerization in eukaryotic membranes. The resemblance between the BiFC and ToxRed assays suggest that biological membranes, despite their eukaryotic or prokaryotic origin, behave similarly, but not equally, when packing TMD helices.

Contrarily to the ToxRED assay, in which the ToxR and MBP moieties have to face the cytosol and the periplasm, respectively, the BiFC approach cannot discern whether the chimeras are being inserted into the membrane or remain cytosolic. To distinguish between these two possibilities, we performed subcellular fractionation treatments in which the membrane and cytosol fractions were separated (Figure 51D). The chimeras bearing 23L, 25L, or wild-type GpA TMDs were located in the membrane fraction (lanes 2, 4, and 6 respectively). Conversely, the EYFP (used as a soluble marker) was found exclusively in the cytosolic fraction (lane 7).



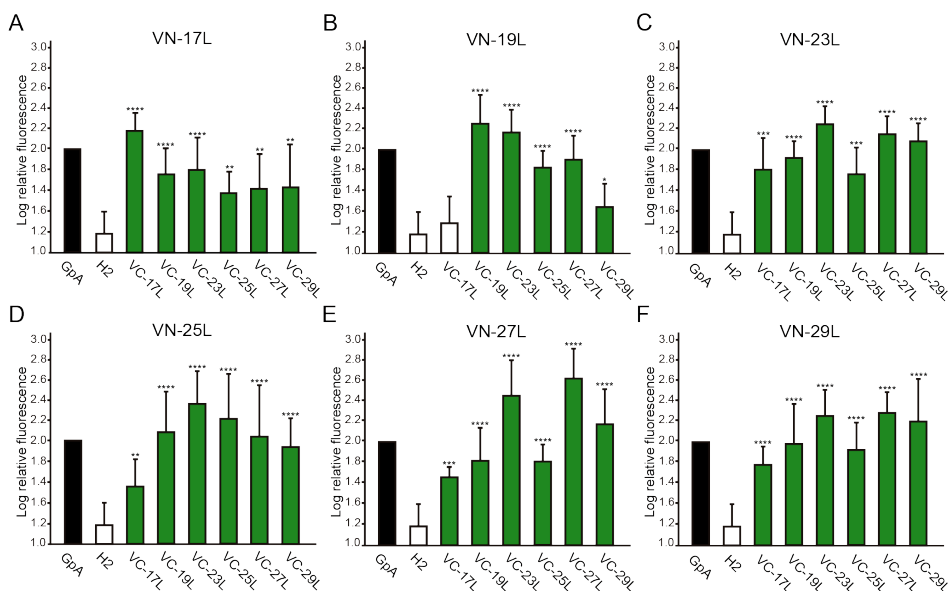
**Figure 51 | Homo-dimerization in eukaryotic membranes.**

(A) Schematic representation of BiFC Assay. TM-driven oligomerization results in complementation of two non-fluorescent halves (amino-terminus (VN) and carboxy-terminus (VC)) of the Venus Fluorescent Protein (VFP). (B) Mean relative fluorescence of chimera homo-oligomerization in HEK-293T cells GpA (VN-GpA/VC-GpA), 17L (VN-17L/VC-17L), 19L (VN-19L/VC-19L), 23L (VN-23L/VC-23L), 25L (VN-25L/VC-25L), 27L (VN-27L/VC-27L), 29L (VN-29L/VC-29L), H2 (VN-H2/VC-H2)). Error bars indicate standard deviation obtained from at least 4 independent replicates (H2 was used as a negative control, \*\*\*\*  $< 0.0001$ ). A color intensity code is used to highlight dimerizing (green) and non-dimerizing (white) transmembrane segments, while positive dimerization control (GpA) is shown in black. (C) Western blot showing chimera's expression levels detected by  $\alpha$ -c-myc antibody. (D) The contribution of Gly for the dimerization of TMD chimeras. Relative fluorescence of chimera homo-oligomerization in human-derived HEK-293T cells (GpA [VN-GpA/VC-GpA (depicted in black)], H2 [VN-H2/VC-H2] (white), 17L [VN-17L/VC-17L], 17L\_I [VN-17L\_I/VC-17L\_I], 29L [VN-29L/VC-29L], 29L\_I [VN-29L\_I/VC-29L\_I]). The bars represent mean values of chimera homo-oligomerization, and error bars denote standard deviation obtained from 3 independent experiments (p-values of Student's t-test: \*  $< 0.05$ , \*\*  $< 0.01$ , ns (non-significant)). Color intensity is used to highlight dimerizing (significantly different from H2 control, green) and non-dimerizing (white). Light green is used to indicate those samples (where the Gly residues of the dimerization domain have been substituted with Ile) whose fluorescence values are significantly higher than the H2 control and at the same time lower than the corresponding non-mutated control. (E) Subcellular fractionation of HEK-293T cells expressing BiFC chimeras (c-myc tagged) (VN-23L, VN-25L, and VN-GpA) or EYFP (Flag-tagged) (used as a soluble marker). Soluble fraction (Sol) and Membrane fraction (MB).

## Results

### c.3.2. Influence of transmembrane hydrophobic length mismatch on heterotypic helix-helix packing

Next, the potential hetero-dimer formation between TMDs with different hydrophobic lengths was analyzed. The GpA homo-dimer was used as a positive reference value ser while the H2 was used as a negative control. Surprisingly, the majority of the tested combinations between VN- and VC- GpA-derived chimeras (all except VN-19L/VC-17L) were capable of reconstituting the VFP and produce fluorescence values significantly higher than the negative control (H2) (Figure 52).

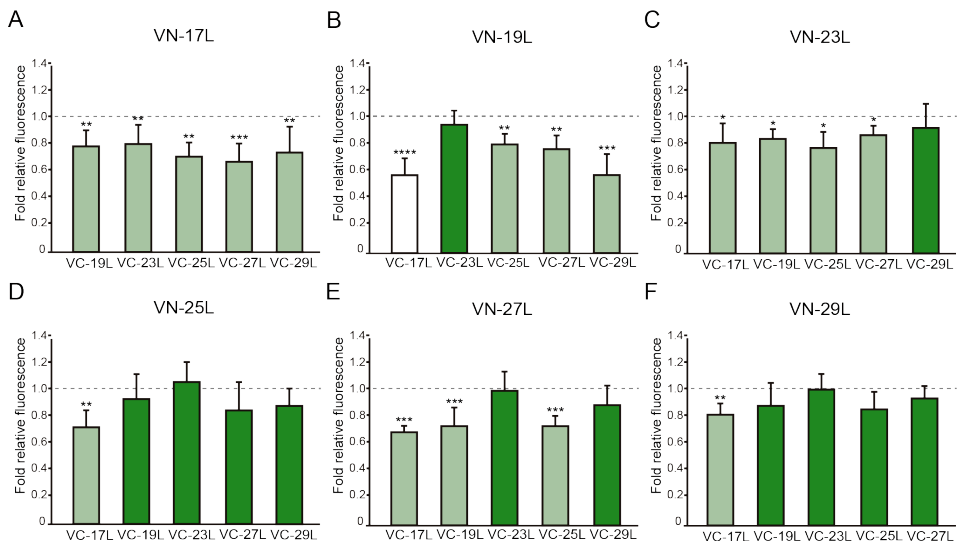


**Figure 52 | Hetero-dimerization in eukaryotic membranes.**

Mean relative fluorescence of chimera hetero-oligomerization in HEK-293T cells of all different combinations: (A) VN-17L/VC-X, (B) VN-19L/VC-X, (C) VN-23L/VC-X, (D) VN-25L/VC-X, (E) VN-27L/VC-X, (F) VN-29L/VC-X. Error bars indicate standard deviation obtained from at least 4 independent experiments. GpA homo-dimer (black bars) was used as positive control and normalization value while Lep H2 homo-oligomer was used as a negative control (white bars). For the experimental samples, a color intensity code was used to highlight dimerization (green, fluorescence values significantly higher than those obtained with the H2 control) and non-dimerization (white, values equivalents to those obtained with the negative control). Additionally, asterisks were included to indicate the level of significance (\*\*<0.01, \*\*\*<0.001, \*\*\*\*<0.0001).

To highlight differences among the tested combinations, data was re-analyzed using the values of the homo-dimers as a reference set. In this case, for any given combination (e.g., VN-X/VC-Y) the corresponding homo-oligomerization values (VN-X/VC-X and VN-Y/VC-Y) were merged and used to obtain the fluorescence fold change unit (Figure 53). Light green bars indicate that the dimerization value for

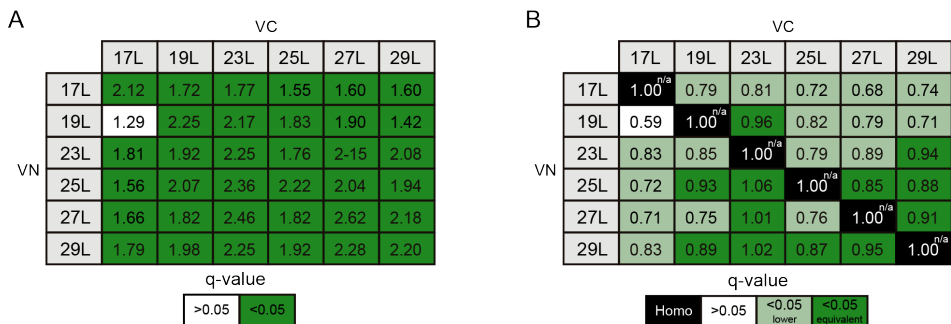
the selected heteromeric combination is significantly lower than its homo-dimer reference ser (the level of significance is depicted with the corresponding number of asterisks on top of each bar). Contrarily, dark green bars show these heteromeric combinations in which the VFP fluorescence was as high as the appropriated homomeric reference set. The VN-19L/VC-17L was depicted in white to indicate that not only the fluorescence was lower than its controls but also not statistically higher than in the non-dimerizing H2 control. Collectively, our data suggest that a different hydrophobic length between the monomers hinders the formation of the dimer. Furthermore, difficulty in hetero-dimer formation in biological membranes can be observed either when a large disparity between the hydrophobic length of the monomers is found or when one of the GpA-derived chimeras contains a hydrophobic region below  $\sim 28$  Å (length of the 17L). Heat map representation of the data in Figure 52 and Figure 53 are included in Figure 54A and B, respectively.



**Figure 53 | Differences in hetero-dimerization in eukaryotic membranes.**

Mean relative fluorescence of chimera hetero-oligomerization. For any given combination (eg. VN-X/VC-Y) the corresponding homo-oligomerization values (VN-X/VC-X and VN-Y/VC-Y) were used as a reference set to obtain fold change and significance (q-values,  $* < 0.05$ ,  $** < 0.01$ ,  $*** < 0.001$ ,  $**** < 0.0001$ ). (A) VN-17L/VC-X, (B) VN-19L/VC-X, (C) VN-23L/VC-X, (D) VN-25L/VC-X, (E) VN-27L/VC-X, (F) VN-29L/VC-X. Error bars indicate standard deviation obtained from at least 4 independent replicates. A color intensity code was used to highlight the dimerization intensity. Dark green (fluorescence values equivalent to the appropriated control), light green (fluorescence values significantly higher than those obtained with the H2 control but significantly lower than those observed with the corresponding homo-oligomer controls), and white (values equivalent to those obtained with the H2 control).

## Results

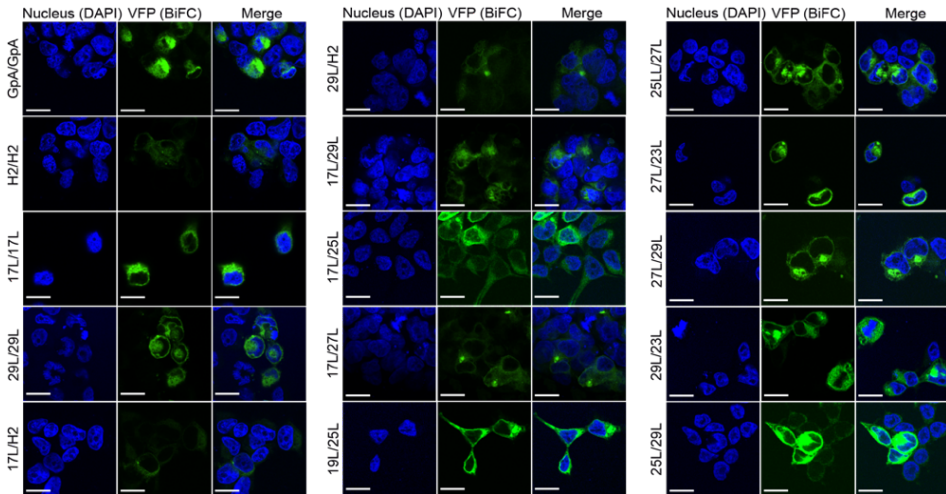


**Figure 54 | Heat map representation of BiFC hetero-oligomerization assay.**

Values represented as bar graphs in Figure 52 and Figure 53. A color intensity code was used to highlight the dimerization intensity as above.

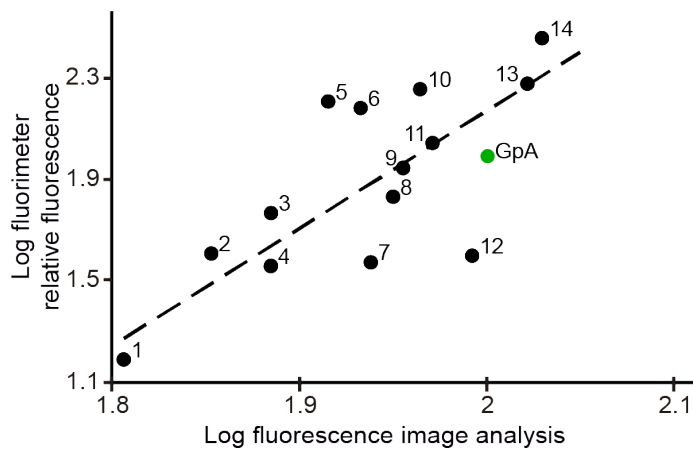
The aforementioned results were corroborated by confocal microscopy (Figure 55). To this end, VN-17L/VC-29L combination was selected as representative of interaction between short-long TMDs. VN-17L/VN-H2 and VN-29L/VC-H2, and the VN-GpA/VC-GpA combinations were used as negative and positive controls. Additionally, the 17L and 29L homo-dimers were included to analyze the behavior of homo-topic dimers. Finally, nine more BiFC combinations (randomly selected) were also included in the assay (VN-17L/VC-25L, VN-27L/VC-27L, VN-17L/VC-29L, VN-19L/VC-25L, VN-23L/VC-27L, VN-25L/VC-27L, VN-25L/VC-29L, VN-27L/VC-29L, and VN-29L/VC-23L) (Figure 55). As in the previous experiment, the confocal images of all tested combinations (including the highly unbalanced VN-17L/VC-29L) but those including H2 showed fluorescence levels above the negative controls. The difference in VFP signal intensity between hetero-dimers that included the 17L TMD and any other oligomer (excluding the negative controls VN-X/VC-H2 or VN-H2/VC-X, here X is any of the tested TMD) were also visible in the fluorescence microscope images (Figure 55). The fluorescence images indicate that the oligomers are located in a perinuclear region, likely the ER membrane. A correlation between fluorescence quantification via fluorescence spectrometry and confocal microscope image analysis is included in Figure 56.





**Figure 55 | Confocal microscopy analysis for membrane dimer formation.**

Confocal microscopy of DAPI stained (blue) HEK-293T cells expressing representative VN/VC combinations (GpA [VN-GpA/VC-GpA], 17L [VN-17L/VC-17L], 29L [VN-29L/VC-29L], 17L/H2 [VN-17L/VC-H2], 29L/H2 [VN-29L/VC-H2], 17L/29L [VN-17L/VC-29L]). Successful TM-driven oligomerization results in VFP reconstitution and fluorescent signal (green). Scale bar size is 16  $\mu$ m.



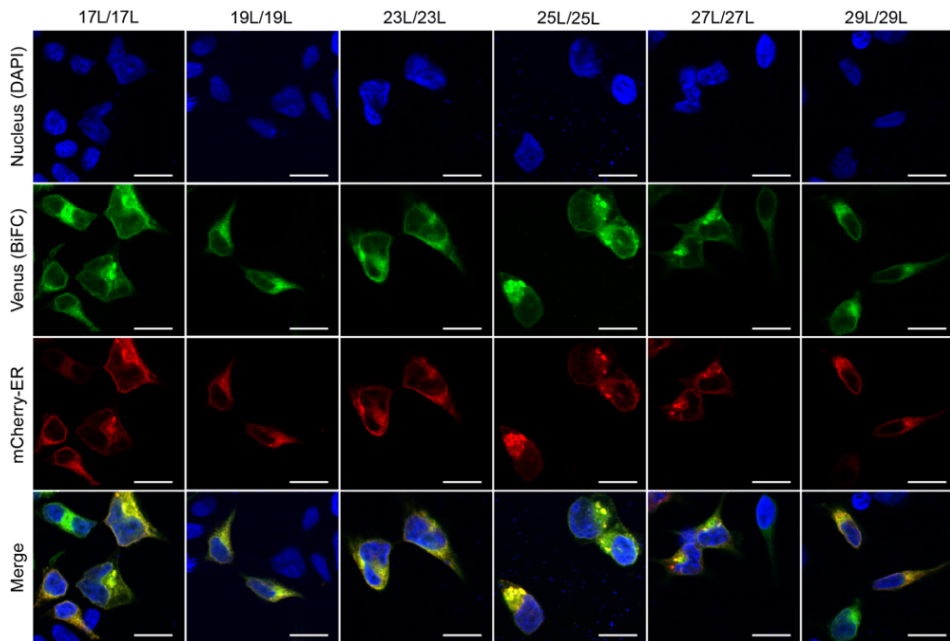
**Figure 56 | Confocal microscopy values vs fluorimeter values.**

Comparison of chimera's VFP reconstitution values obtained via fluorimeter or confocal images quantification. The dots correspond to the combination included in Figure 55 (1 (VN-17L/VC-H2), 2 (VN-17L/VC-29L), 3 (VN-17L/VC-25L), 4 (VN-23L/VC-25L), 5 (VN-29L/VC-29L), 6 (VN-27L/VC-29L), 7 (VN-29L/VC-H2), 8 (VN-19L/VC-25L), 9 (VN-25L/VC-29L), 10 (VN-29L/VC-23L), 11 (VN-25L/VC-27L), 12 (VN-17L/VC-27L), 13 (VN-17L/VC-17L), 14 (VN-27L/VC-23L)). GpA was highlighted in green (not included in the trend line analysis).

## Results

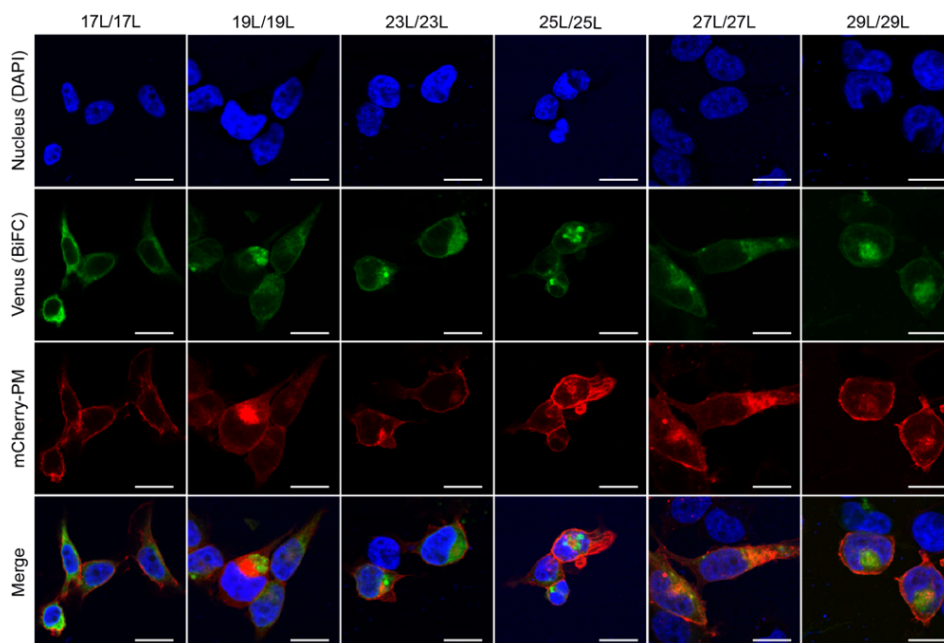
### c.3.3. Influence of hydrophobic length on the subcellular localization of the chimeras

The influence of the hydrophobic length on the subcellular localization of single spanning TMD has been reported (Sharpe et al. 2010; Cosson et al. 2013). However, previous confocal microscopy results (Figure 55) suggested a perinuclear localization of most chimeras, regardless of their hydrophobic length. To investigate the subcellular localization of the homo-oligomers VN- VC- corresponding plasmids together with ER (Figure 57) or plasma membrane (PM) (Figure 58) marker were co-expressed in HEK-293T cells.



**Figure 57 | Homo-dimerization at the endoplasmic reticulum.**

Confocal microscopy of DAPI stained (blue) HEK-293T cells expressing tested homo-dimers (17L [VN-17L/VC-17L], 19L [VN-19L/VC-19L], 23L [VN-23L/VC-23L], 25L [VN-25L/VC-25L], 27L [VN-27L/VC-27L], 29L [VN-29L/VC-29L]). TM-driven homo-oligomerization results in VFP reconstitution and fluorescent signal (green). Sec61 $\beta$  fused to mCherry fluorescent protein was used as ER marker (red). The presence of colocalization of red and green signals in the merge images was highlighted in yellow. Scale bar size was set to 16  $\mu$ m.

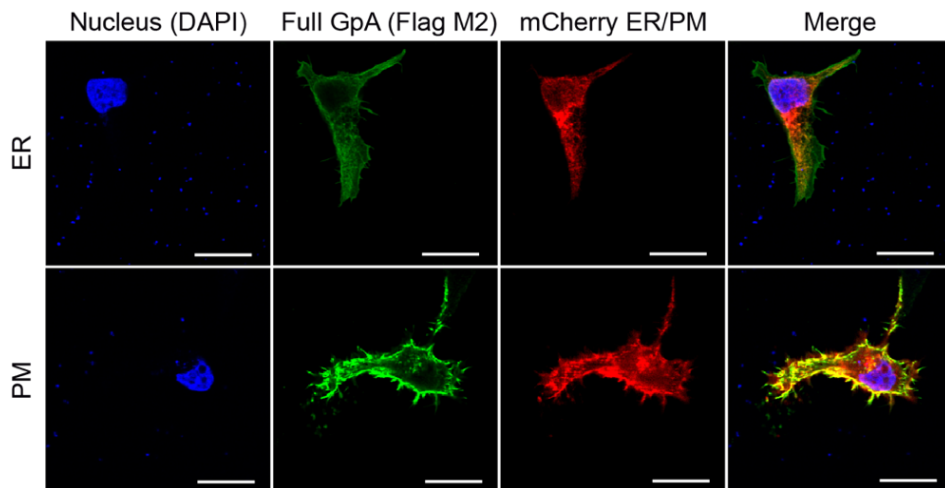


**Figure 58 | Homo-dimerization at the plasma membrane.**

Confocal microscopy of DAPI stained (blue) HEK-293T cells expressing tested homo-dimers (17L [VN-17L/VC-17L], 19L [VN-19L/VC-19L], 23L [VN-23L/VC-23L], 25L [VN-25L/VC-25L], 27L [VN-27L/VC-27L], 29L [VN-29L/VC-29L]). TM-driven homo-oligomerization results in VFP reconstitution and fluorescent signal (green). Neuromodulin fused to mCherry fluorescent protein was used as PM marker (red). The presence of colocalization of red and green signals in the merge images was highlighted in yellow. Scale bar size was set to 16  $\mu\text{m}$ .

The microscope-based assay revealed no considerable differences in subcellular localization among the tested TMD and indicate that all the homo-oligomers tested, regardless of their hydrophobic length, are preferentially found in the ER membranes. Being our chimeras based on the GpA TMD, it was surprising to find them in the ER membranes and not in the PM (Pang & Reithmeier 2009). To confirm that the results were not a mere artifact, HEK-293T cells were transfected with a plasmid bearing the full sequence of GpA (Flag tagged) and analyzed its co-localization with ER and PM markers. The micrographs clearly indicate that the full-length GpA localizes in the PM membrane (Figure 59). Therefore, since the subcellular localization of all TMD studied is similar we can assume that the hydrophobic match between the hydrophobic length of the TMD and the membrane stands as a major contributor to the differences observed in the present TMD packing study.

## Results



**Figure 59 | Sub-cellular localization of full-length GpA.**

Confocal microscopy of DAPI stained (blue) HEK-293T cells expressing GpA full-length protein (Flag-tagged). Neuromodulin fused to mCherry fluorescent protein was used as a plasma membrane marker (PM) while Sec61 $\beta$  fused to mCherry fluorescent protein was used as endoplasmic reticulum marker (ER).

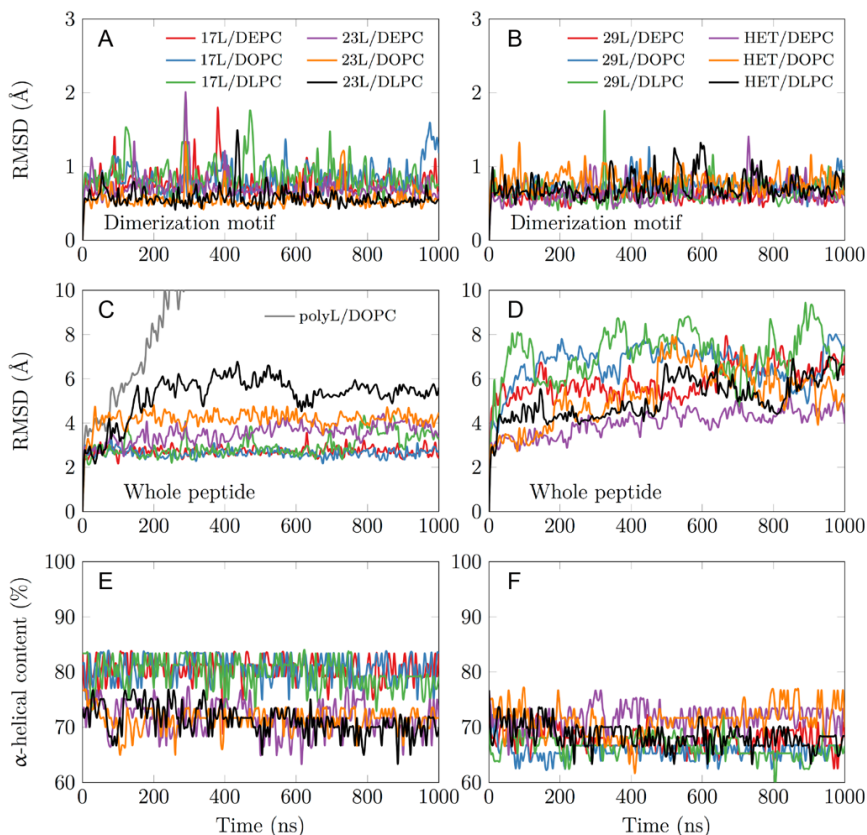
### c.3.4. Molecular dynamics simulation of homo- and hetero-oligomerization in different size membranes

To gain more detailed insight into how cellular membranes adapt to hydrophobic mismatch, atomistic MD simulations were performed by the Vattulainen's Lab from the University of Helsinki (<https://www.helsinki.fi/en/researchgroups/biophysics>). Initially, 17L, 23L and 29L TMDs (containing the minimized dimerization motif) were embedded in single-component DLPC (12:0 PC), DOPC (18:1 PC), and DEPC (22:1 PC) bilayers. This choice of bilayer systems provides a systematic increase in membrane thickness. The hydrophobic thickness values of 20.0 Å, 27.4 Å, and 35.7 Å correspond to phosphorous-phosphorous thicknesses of 31.6 Å, 38.7 Å, and 47.3 Å, respectively. The DOPC bilayer is likely the closest mimic of the membranes studied *in vivo*, as palmitic (16:0), palmitoleic (16:1), and oleic (18:1) fatty acids are the most common lipid chains in *E. coli* (De Siervo 1969) and HEK-293T (Dawaliby et al. 2016) membranes. Together with the hydrophobic lengths of 25.5 Å, 34.5 Å, and 43.5 Å estimated for the 17L, 23L, and 29L TMDs, respectively (Table 1), the different combinations allow to consider both positive (TMD hydrophobic length > membrane thickness) and negative (TMD hydrophobic length < membrane thickness) mismatch. Furthermore, the unsaturation of the longer chains ensured that

all bilayers remain in the liquid disordered phase at the physiological temperature of 37 °C.

All GpA-based dimers maintained their TMD positioning in the membrane and were stable during the 1  $\mu$ s simulations. This is evidenced by the time evolution of the RMSD of the dimerization motif (defined as GVxxGVxxT) shown in Figure 60A and B, which suggest that this region is equally stable regardless of the mismatch. This observation corroborates the experimental findings described above, showing that all GpA chimeras containing the minimized dimerization motif were capable to form stable homo-dimers. As shown in Figure 60C, the polyleucine (lacking the dimerization motif) dimer dissolved rapidly in a DOPC membrane. The RMSD of all the rest of GpA-based dimers reveals that the shortest 17L dimmers were overall very stable due to their location within the membrane (Figure 60A and B), while the ends of the longest 29L dimers were more mobile since they reside in the aqueous phase. The stability of the 23L dimers decreases upon increasing membrane thickness as their ends become more exposed to water and therefore more prone for unfolding. The increase in the fluctuations within the termini region due to increasing mismatch (negative to positive) was also evident in the RMSF data shown in and explained by a lower overall  $\alpha$ -helical content of the peptides (see Figure 60E and F, and Figure 61). Interestingly, the  $\alpha$ -helical content of the peptides in a given dimer is independent of the thickness of the host membrane. However, when placed in the same membrane, the longer the peptides are, the lower their  $\alpha$ -helical content is. Still, although the percentages of the  $\alpha$ -helical content decrease as the peptides get longer, the absolute number of the amino acids in the  $\alpha$ -helical conformation actually increases. Thus, some but not all added Leu residues, originally in a helical conformation, unfold rapidly during the short equilibration simulations, during which position restraints are turned off.

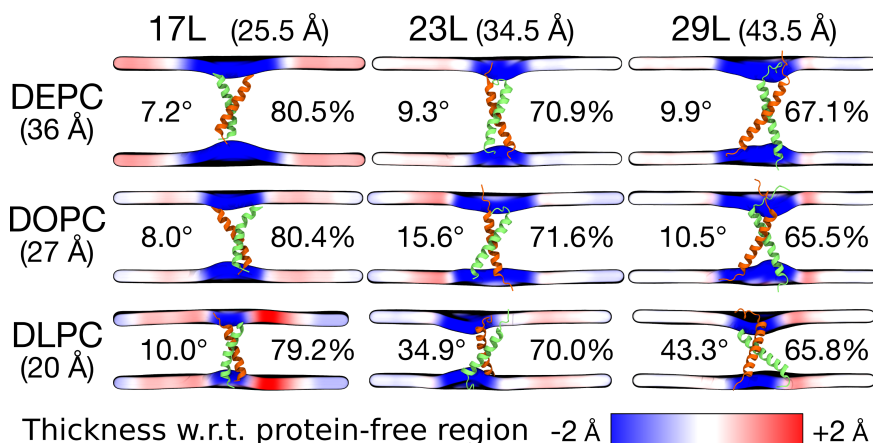
## Results



**Figure 60 | Stability of the dimeric structures from molecular dynamics simulations.**

(A-D) Root mean squared deviation (RMSD) of the dimerization motif (5 residues per peptide) is shown on the top row (A, B) and the RMSD of the whole dimer on the bottom row (C, D). The whole trajectory is included in the analyses. (E, F) Helicity of the peptides during the full simulation time period. The hetero-dimer is labeled ‘HET’ and the polyleucine control in DOPC as ‘PolyL/DOPC’.

Despite the varying levels of mismatch, the stability of the dimers suggests that either the peptide dimer or the membrane is able to compensate for effects induced by the mismatch. The compensation likely takes place through structural adjustment. In the case of positive mismatch, the tilt angle of the dimer (Figure 61 and Figure 62) reveals that the collective tilting of the dimer plays a role only in the case of the longer peptides (23L and 29L) simulated in the thinnest DLPC membrane, where the mismatch is larger, causing the dimer to tilt significantly ( $\sim 40^\circ$ ) maintaining the crossing angle between monomers. In other cases, the dimer stands almost upright in the membrane (average tilt angle being  $< 20^\circ$ ). Meanwhile, in the case of a negative mismatch, the tilt angle remains low. Notably, there is little adaptation to the hydrophobic mismatch by a scissor-like motion of the dimers, as seen in the peptide crossing angles plots (Figure 62A and B).

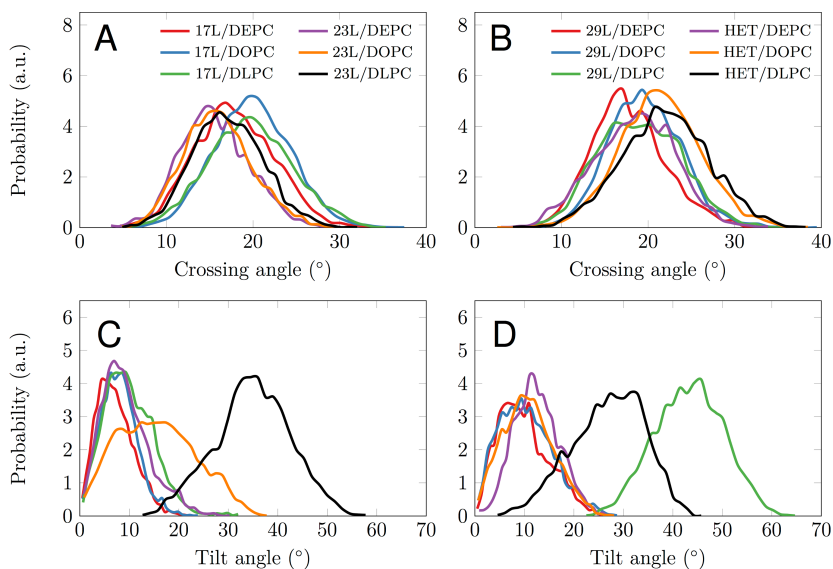


**Figure 61 | Summary of the results from molecular dynamics simulations of homo-dimers.**

The peptide and the membrane are unchanged along the columns and rows, respectively. The hydrophobic length of the peptide and the hydrophobic thickness of the membrane are given in brackets. The two peptides are shown in green and orange, while the mean positions of the phosphorus atoms in lipid head groups are shown as a surface. This surface is colored according to the value of local average membrane (phosphorus-phosphorus) thickness with respect to the bulk membrane thickness far away from the dimer. The thicknesses are calculated from the last 500 ns of the simulations. The average tilt of the peptides is given in degrees and the average  $\alpha$ -helical content in percentage.

For both positive and negative mismatch, the membrane thickness was perturbed only locally as demonstrated in Figure 61. The spatial extent of the perturbation depends on the level of mismatch, though. For the 17L dimer, the effect was smallest in the case of DLPC membrane and increases systematically with increasing bilayer thickness (DLPC < DOPC < DEPC). For the 23L and 29L dimers, the smallest perturbation was observed in the case of DEPC membrane. The effect became more prominent in the DOPC membrane, which is thinner than DEPC, however, the effect did not increase further in the DLPC membrane as the tilting of the dimers began to dominate the membrane organization. Notably, the differences in the behavior of the studied peptides in the DOPC membrane, whose thickness resembles that of the bilayers studied here *in vivo*, are insignificant.

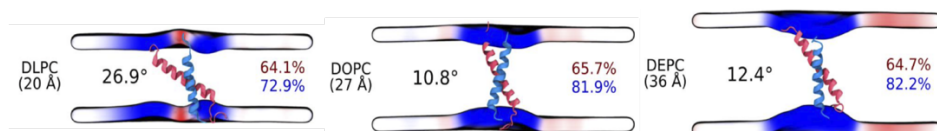
## Results



**Figure 62 | Tilt and crossing angle distribution.**

(A, B) Crossing angle distribution of the dimers calculated from the last 500 ns of the simulations. (C, D) Tilt angle distribution of the dimers calculated from the last 500 ns of the simulations. The 17/29L hetero-dimer is labeled ‘HET’.

As a result of homo-dimer simulation, computer simulations of hetero-dimer with the biggest differences in monomer length (17L/29L) were performed. The 17L/29L hetero-dimer also remains stable and inside the membrane (Figure 60B and D). Similarly to the 23L and 29L homo-dimers, the main adaptation mechanism for the 17L/29L hetero-dimer in DEPC and DOPC membranes was membrane thickness perturbation, while in the DLPC membrane the whole dimer tilted significantly, maintaining the overall crossing angle between monomers (Figure 63). Furthermore, the  $\alpha$ -helical content of the individual peptides forming the hetero-dimer was similar to their values in the corresponding homo-dimers, except for the 17L peptide of the hetero-dimer in the DLPC membrane, for which this value was somewhat lower than in the homo-dimer.



**Figure 63 | Summary of the results from molecular dynamics simulations of the 17L/29L hetero-dimer.**

29L (43.5 Å) peptide is red colored while the 17L (25.5 Å) peptide is blue colored. Membrane coloring and organization as in Figure 61, except that the  $\alpha$ -helical contents are given separately for the two peptides, colored as the helix.



## Chapter 4. Topological study of MPs

MPs are involved in diverse processes, featuring important functions such as receptors and transporters across the membrane, immune system molecule recognition, energy transduction, protein secretion and folding, viral recognition and penetration, etc. Thus, understanding their mechanisms of folding, insertion, or topology is of great importance to continue increasing our knowledge about how cells work.

The topology of a MP describes the number and approximate locations of the TMDs on a given protein, as well as the overall orientation (locations of N- and C-termini) in a membrane. In the present chapter, we are unraveling the topology of some relevant MPs in order to better describe the processes in which they are involved.

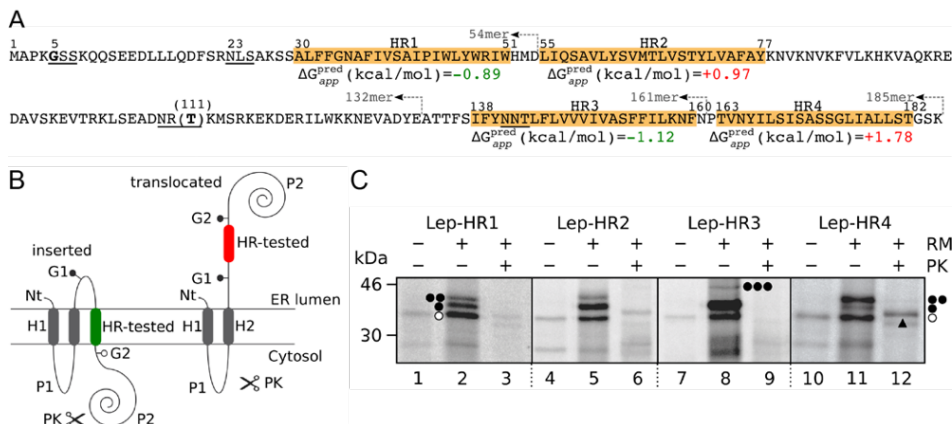
### c.4.1. Membrane topology of the translocon-associated protein gamma subunit (TRAP $\gamma$ )

Despite notorious efforts, the role of the TRAP complex during membrane insertion or translocation is not fully understood. Nevertheless, it has been proposed that the complex acts facilitating the initiation of protein translocation, as a membrane protein regulator (G. von Heijne 1984; Fons et al. 2003), among other functions. The TRAP $\gamma$  protein is the largest subunit of the TRAP complex (i.2.3.3). The gamma subunit assumes a central position, binding the ribosome on the cytosol face of the ER membrane and coordinating the remaining TRAP subunits with the ribosome and other translocon components. Although its importance for secreted and membrane proteins biogenesis has been proved (Sommer et al. 2013; Nagasawa et al. 2007), its membrane topology and biogenesis have not been thoroughly investigated.

For this purpose, identification and location of the TRAP $\gamma$  subunit TMDs were done by using the algorithm  $\Delta$ G Prediction Server v1.0 (<http://dgpred.cbr.su.se>) (Hessa et al. 2005; Hessa et al. 2007) (Figure 64A). Once the algorithm predicted the TMDs (hydrophobic residues 30-51 (HR1), 55-77 (HR2), 138-160 (HR3), 163-182 (HR4)), they were cloned in the Lep system in order to determine their individual membrane insertion capabilities (Figure 64B). All segments resulted mainly in single-glycosylated forms (HR1, HR2, and HR3) except for HR4 (Figure 64B, lane 11) which was found mainly double-glycosylated. It should be mentioned that TRAP $\gamma$  sequence includes a native potential N-glycosylation site at position 141, and the translocated form in the Lep-HR3 is represented by triple glycosylation since its native glycosylation site is also available for the OST catalytic site when the HR3 is

## Results

not inserted (Figure 64B, lane 8). As expected, digestions with proteinase K of Lep chimeras bearing HR1, HR2, and HR3 were sensitive to the treatment whereas the construct containing HR4 sequence was partially resistant to the protease treatment due to its luminal P2 location.



**Figure 64 | Insertion of predicted TMDs into microsomal membranes.**

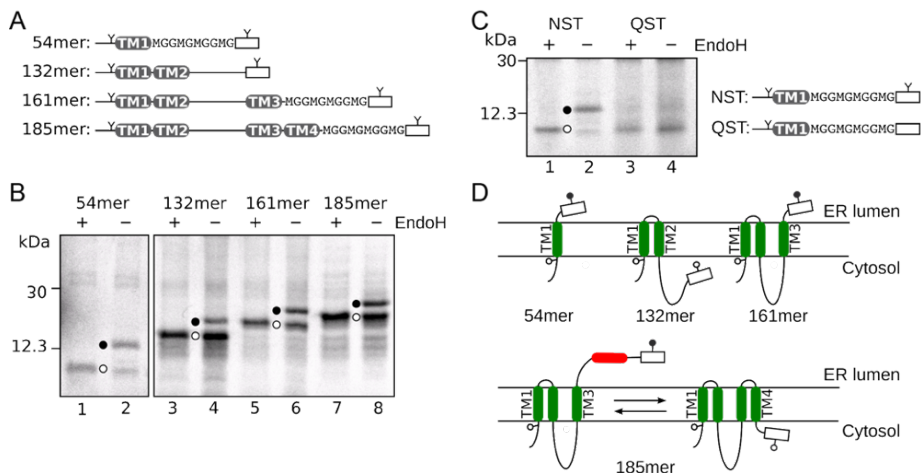
(A) Amino acid sequence of TRAP $\gamma$ . Predicted TMDs by the  $\Delta G$  algorithm (<http://dgpred.cbr.su.se>) are highlighted in boxes and their predicted apparent values are shown at the bottom. Negative  $\Delta G_{app}$  values, indicative of insertion, are shown in green. Positive  $\Delta G_{app}$  values, indicative of membrane translocation, are shown in red. Glycosylation acceptor sites naturally find in the sequence are underlined. A set of oligonucleotides (arrowhead dotted lines) was designed to generate TRAP $\gamma$  truncated forms, which are indicated by the length of the truncated polypeptide (ending residue -mer). (B) Schematic representation of the Lep system. (C) *In vitro* translation of the different Lep constructs bearing the HR1 (residues 30-51), HR2 (residues 55-77), HR3 (residues 138-160), HR4 (residues 163-182) in the presence (+) and absence (-) of rough microsomes (RM) and proteinase K (PK). Bands of non-glycosylated proteins are indicated by a white dot; single and doubly glycosylated proteins are indicated by one and two black dots, respectively. In the case of Lep-HR3 construct a triple glycosylated band ( $\approx 2\%$ ) was observed (lane 8) due to the presence of an acceptor NT (residues 141-143) within the translocated hydrophobic region. The protected doubly-glycosylated H2/G1/HR4/G2/P2 fragment is indicated by an arrowhead.

To experimentally map the membrane topology of TRAP $\gamma$  subunit, we prepared a series of polypeptide truncates containing an added C-terminal glycosylation tag (NST). The mapping was done through a series of polypeptide truncated containing an added C-terminal glycosylation tag (NST), which has proven to be efficiently modified in the *in vitro* coupled transcription and translation in presence of microsomes (Doñate-Macian et al. 2015; Bañó-Polo et al. 2011). Since it has been previously reported that glycosylation occurs when the acceptor Asn is at least  $\sim 11$ -13 residues away from the membrane (Bañó-Polo et al. 2011; Nilsson & G. von Heijne 1993; Pfeiffer et al. 2014), the C-terminal glycosylation tag where placed spaced as widely as possible from the TM. For TMD without enough space until the next TMD (HR1, HR3, and HR4) a glycosylation tag with a Met and Gly hanging to

separate the glycosylation acceptor site at least 13 residues from the end of the TMD. Truncated constructs of the TRAP $\gamma$  sequence with the NST tag are shown in (Figure 65A). Translation products containing the N-terminal 54 residues of TRAP $\gamma$  sequence, including the first predicted TM1 segment plus an optimized glycosylatable C-terminal tag (54mer truncate) were efficiently singly-glycosylated in the presence of microsomal membranes (Figure 65B, lane 2). Nonetheless, since wild-type TRAP $\gamma$  subunit carries a native potential glycosylation site at position 23 (see Figure 64A) with theoretically not enough distance to the predicted start of TM1 (position 30) to be glycosylated, a 54mer translation with a non-glycosylatable tag was performed to corroborate the results. The efficient glycosylation observed only when using the C-terminal acceptor site NST (Figure 65C, lane 2) in comparison with the non-acceptor site QST (Figure 65C, lane 3) strongly indicate that TM1 is properly recognized by the translocon machinery to be inserted into the membrane with an N-terminal cytoplasmic orientation (see Figure 65D for a scheme).

Truncated polypeptide 132mer were poorly glycosylated (Figure 65B, lane 4), indicating that the second predicted TM2 segment efficiently integrates into the membrane (see Figure 65D for a scheme). The insertion of the TM3 predicted segment was tested by translating a 161mer polypeptide with the same glycosylatable tag used with 54mer polypeptide. The glycosylation of this truncated (Figure 65B, lane 6) indicated that TM3 was inserted with its C-terminus oriented towards the ER lumen (see Figure 65D for a scheme). Finally, the intermediate glycosylation efficiency observed for full-length protein (185mer) translations (Figure 65B, lane 8) supports the existence of an equilibrium between insertion-translocation for this last TMD, suggesting that the presence of preceding TMDs causes a noticeable increase in the insertion efficiency of HR4, but not enough for full assembly into the microsomal membranes, as illustrated in Figure 65D (bottom). Subsequently, experiments performed in mammalian cells (data not shown) supported the full-length protein topology acquisition after translating the full polypeptide. It seems that the TRAP $\gamma$  subunit final topology acquisition requires the presence of some additional cellular components to get the protein native conformation.

## Results



**Figure 65 | *In vitro* analysis of truncated TRAP $\gamma$  constructs.**

(A) Schematic representation of the constructs used in the assay. To monitor the membrane orientation of truncated TRAP $\gamma$  molecules a glycosylatable (NSTMSM) tag (white rectangle) was added at position 54 (54mer), 132 (132mer), 161 (161mer), and 184 (184mer). The position of the glycosylation sites is marked with a Y symbol. The presence of a TMD identified by the  $\Delta$ G Prediction Server (<http://dgpred.cbr.su.se>) in each construct and the required linker sequence preceding the glycosylatable tag to allow glycosylation is also included in 54mer, 161mer, and 185mer. (B) *In vitro* translation of the 54mer, 132mer, 161mer, and 185mer truncates in the presence of rough microsomes (RM). Lanes 1, 3, 5, and 7, samples were treated with endoglycosidase H (EndoH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while a black dot indicates glycosylation of the C-terminal tag. Glycosylation percentages are shown at the bottom. (C) *In vitro* translation in the presence of RM of 54mer truncated constructs bearing an acceptor (NST, NSTMSM) or non-acceptor (QST, QSTMM) C-terminal glycosylation tag. White and black dots indicate non-glycosylated and glycosylated molecules respectively, as in panel B. (D) Schematic representation of the membrane topology of the 54mer, 132mer, 161mer and 185mer truncates. A hydrophobic region is noted as a green box when inserted in the membrane, or as a red box if it is not recognized by the translocon as a TMD. The position of the glycosylatable tag (white rectangle) and its glycosylation status (white and black dots) are also shown. Two potential topologies for the 185mer are represented (bottom). Experiments were done and quantified at least in triplicate.

### c.4.2. Membrane topology of bacteriorhodopsin

MPs in general, and rhodopsin-like G protein-coupled receptors (GPCRs) in particular, are of significant biological and medical relevance since they represent over 50% of current drug targets (Hilger et al. 2018). These GPCRs are seven helix membrane proteins, which receive an optical or chemical signal on the extracellular membrane surface and initiate G-protein based signaling cascades in the cytosol. In this subset of MPs one of the most studied proteins is bacteriorhodopsin (bR), an

integral MP found in the cytoplasmic membrane of *Halobacterium salinarum*. The apoprotein of bR is structurally well-characterized with an N<sub>out</sub>/C<sub>in</sub> topology and containing seven TMD  $\alpha$ -helices (named A to G). The atomic structure of bR has been described as a trimer in which each monomer is formed by a heptahelical bundle with a retinal chromophore in the polypeptide core (Luecke et al. 1999).

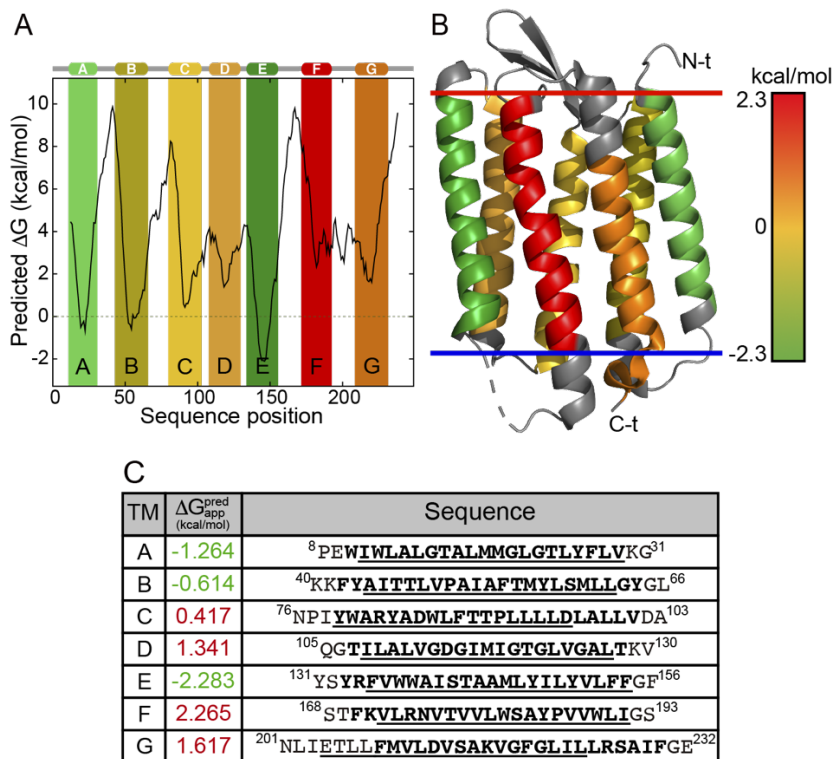
The study of bR has been essential for understanding how MPs fold e.g., interactions stabilizing individual structural elements such TMDs and polypeptide loops were established using bR as a model protein (Müller et al. 2002; Joh et al. 2008). Furthermore, most of the methods to probe helical MP folding mechanisms, including kinetic, thermodynamic, and mechanical approaches to monitor folding and unfolding equilibria were also originally developed using bR (Popot et al. 1987; Booth et al. 1995; Janovjak et al. 2003; Faham et al. 2004). In this direction, bR folding appears to proceed around a helical core in a remarkably cooperative manner (Dale et al. 2000; Booth 2012), in clear contrast to the distinct, processive helix formation and packing stages predicted by the two-stage model (Popot et al. 1987; Popot & Engelman 1990). Previous studies have demonstrated that some helices are independently formed but others are stabilized by tertiary contacts (Hunt et al. 1997; Marti 1998; Krishnamani & Lanyi 2012), resulting in cooperative secondary and tertiary structure formation.

Here, an *in vitro* system (OST glycosylation assay) was used to study the biogenesis of bR in the presence of microsomal membranes. Firstly, identification and location of the bR TMDs were done by using the algorithm  $\Delta G$  Prediction Server v1.0 (<http://dgpred.cbr.su.se>) (Hessa et al. 2005; Hessa et al. 2007) (Figure 66A). The algorithm predicted proper insertion in the case of helices A, B, and E, where a negative  $\Delta G$  was estimated (Figure 66C). However, low insertion capabilities for helices C, D, F, and G were predicted according to their positive  $\Delta G$  values. Surprisingly, the predicted TMDs do not fully matched with the TMDs annotated in the bR structure deposited in the PDB database (PDB code: 1C3W) (Figure 66B).

Bearing this biochemical and biophysical information in mind, truncated polypeptide chains of increasing length from the bR sequence with a C-terminus glycosylation tag (C-tag) as a topological marker was designed (see Annex VI for complete bR sequence). Since no signal sequence was included in our bR-derived sequences, an extra N-terminus glycosylation tag (N-tag) was included as an extra topological marker. N-tag contains the 13 first amino acids of the Lep' model protein, harboring an optimized glycosylation acceptor site (Lundin et al. 2008). In the case of the predicted helix A, four different truncated versions were designed, with (NST) or without (QST) glycosylatable tags (41mer truncate). These 4 truncated sequences contain either two acceptor sites in both N- and C-termini, or one only in N- or C-terminus, or without acceptor site (Figure 67A). The glycosylation profile of the

## Results

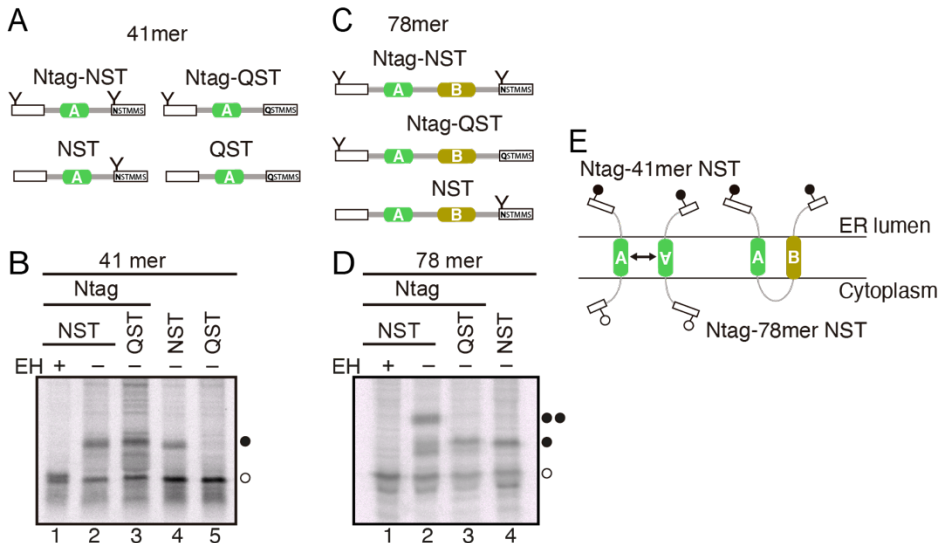
41mer truncate showed a singly-glycosylated band independently of the acceptor site location (Figure 67B, lanes 2, 3, and 4). These results suggested a dual topology for the insertion of the first TMD (41mer) (Figure 67E, left).



**Figure 66 | Predicted Bacteriorhodopsin TMD by  $\Delta G$  Predictor server fitted in bR structure.** (A) Full protein screening of the predicted  $\Delta G$  for the insertion of hydrophobic regions (calculated by the  $\Delta G$  prediction server v1.0). Black line is the  $\Delta G$  prediction for a sliding window of 21 amino acids. Predicted TMDs are highlighted in colored bars according to their  $\Delta G$  values (green to red from most to least hydrophobic TMD). (B) bR structure (PDB code: 1C3W) with orientation  $N_{out}/C_{in}$ . Helices are colored as (A). Red (extracellular) and blue (cytoplasmic) lines indicate approximate membrane location calculated using the PPM server (M. A. Lomize et al. 2012) (C) Comparative table between the predicted (bold amino acids) and PDB annotated (underlined amino acids) TMDs. The annotated TMDs were extracted from the structure deposited in the PDB database (PDB code: 1C3W). Predicted apparent free energy values are shown in red (positive values, indicative of no-insertion) and green (negative values, indicative of insertion).

Insertion of the second predicted TMD was tested by translating a 78mer truncate with the same glycosylation tags as reporters (Figure 67C). Glycosylation profile of the truncated polypeptides showed double-glycosylated molecules (Figure 67C, lane 2) associated with an  $N_{out}/C_{out}$  topology (Figure 67E, right). The detection of a doubly (Figure 67C, lane 2) and singly-glycosylated bands (Figure 67C, lane 3) revealed a TMD AB  $N_{out}$  predominant topology. Interestingly, previous refolding

experiments suggested proper insertion of the AB helical hairpin in lipid vesicles (Kahn & Engelman 1992).



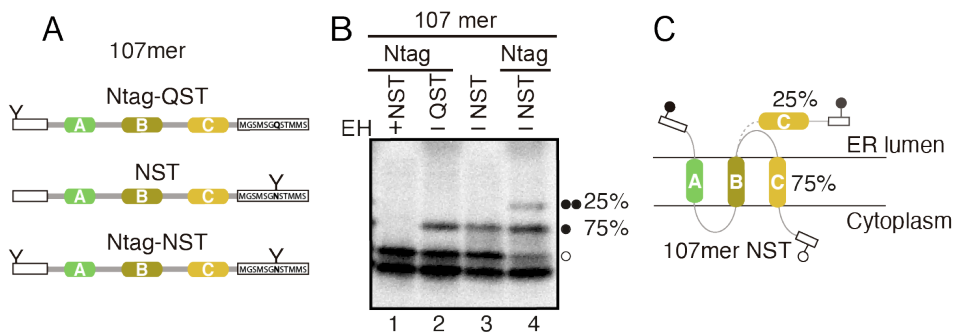
**Figure 67 | *In vitro* expression and representative SDS-PAGE analysis of short bR truncates.**

(A, C) Schematic representation of the TMD A truncates used in the assay. TMDs are colored as in Figure 66A. To identify the membrane orientation of truncated bR molecules a glycosylatable (NST, **NSTMSM**) tag (letter containing white rectangle with a Y symbol) was added at position 41 (41mer) (A) and 78 (78mer) (C). Non-acceptor (QST, **QSTMSM**) tag (white rectangle) was placed to ensure a correct topology characterization. (B, D) *In vitro* translation of the 41mer (B) and 78mer (D) truncated variants in the presence of microsomes. Lane 1, sample was treated with endoglycosidase H (EH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while a black dot indicates glycosylation of the tag. (E) Schematic representation of the membrane topology of the 41mer and 78mer truncates. The position of the glycosylatable tag and its glycosylation status (white and black dots) is also shown. The 41mer truncated is able to adopt both topologies,  $N_{out}$  and  $N_{in}$ .

To test the insertion of the third TMD (helix C), polypeptide truncated versions up to residue 107 were designed (Figure 68A). Those 107mer truncated molecules were C-terminally extended with an amino acid linker that ensures proper glycosylation distance from the membrane (Braunger et al. 2018). The linker is a requirement since, as mentioned before, the MGD between the end of a TMD and the acceptor site is 12/13 residues (Nilsson & G. von Heijne 1993), and in this case there are only five residues between helices C and D (see Figure 66A and Annex VI). The amino acid sequence for the extended tag was MGSMSGNSTMMS.

## Results

Alternatively, the non-acceptor tag MGSMSGQSTMMS was used as a control. Translation/insertion *in vitro* of 107mer-NST truncate molecules in the presence of RMs (Figure 68B, lane 4) renders a significant single-glycosylated form (75% of the total glycosylated molecules). The presence of a light double-glycosylated band (Figure 68B, lane 4) suggested a non-efficient insertion of helix C and arises from an equilibrium between the inserted and translocated forms (Figure 68C), as recently reported in a reductionist approach in which the insertion of helix C was studied out of its native context (Bañó-Polo et al. 2020).



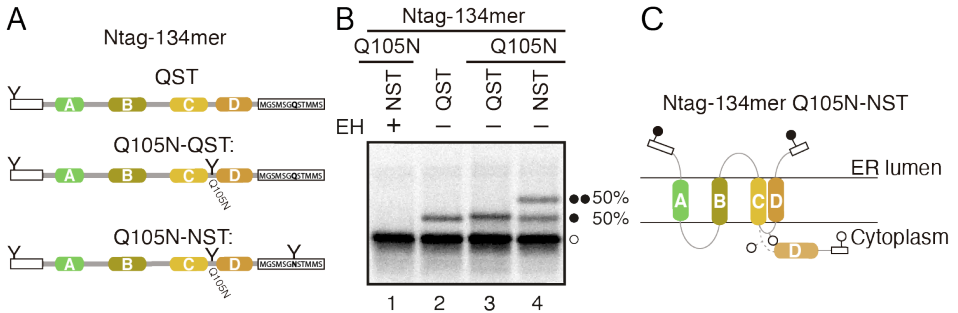
**Figure 68 | Glycosylation mapping of ABC truncated version of bacteriorhodopsin.**

(A) Schematic representation of the TMD C truncated (107mer) with (NST, MGSMSGNSTMSM) or without (QST, MGSMSGQSTMMS) extended glycosylatable tag represented as a letter containing white rectangle with a Y symbol or just as a white rectangle, respectively. TMDs are colored as in Figure 66A. (B) *In vitro* translation of the 107mer truncated variants in the presence of rough microsomes. Lane 1, sample was treated with endoglycosidase H (EH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while one or two black dots indicates glycosylation of one or two glycosylation tags, respectively. (C) Schematic representation of the membrane topology of the 107mer truncate. The position of the glycosylatable tag and its glycosylation status (white and black dots) is also shown. Percentage of singly- and doubly-glycosylated bands are calculated as the fraction of the total glycosylated bands. Two potential topologies for the 107mer and its abundances are represented.

Next, the construct that contains the first four helices of bR (134mer) was studied (Figure 69A). As for truncated ABC, an extension was also required since the loop between helices D and E includes only four residues (see Figure 66A and Annex VI). To determine the correct topology, an extra acceptor site was placed in the loop between helices C and D (position Gln105), a conservative mutation that creates 134mer Q105N truncates (Figure 69A). The need for an extra acceptor site is caused by the helix C non-efficient insertion, which may complicate the proper topology determination if cannot be discerned between insertion or translocation of both helices C and D (without the Q105N acceptor site both possibilities will render double-glycosylated bands). The translation of the construct 134mer Q105N-NST rendered mainly double glycosylated forms (Figure 69B, lane 4), confirming that TMD C and TMD D are inserted together as did not occur for 107mer where TMD



C remained partially translocated when translated in the absence of the following helix D. When a non-acceptor site was used (134mer Q105N+QST), a singly glycosylated form was observed (Figure 69B, lane 3), consistent with the insertion of the polypeptide chain by helices A-D.

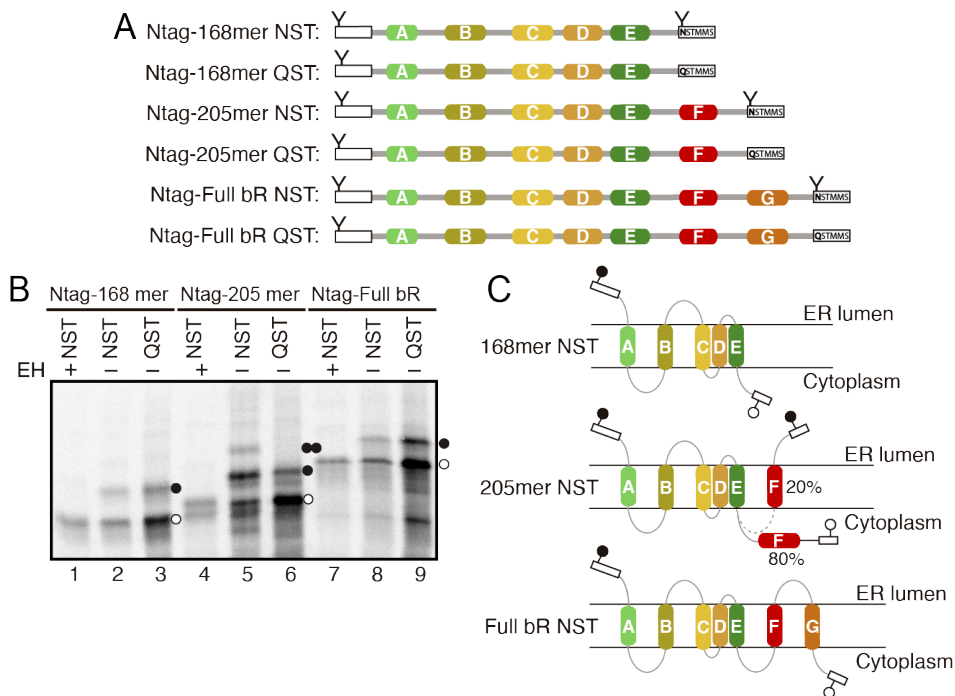


**Figure 69 | Glycosylation mapping of A-D truncates of bacteriorhodopsin.**

(A) Schematic representation of the TMD D truncated (134mer) with (NST, MGSMMSGNSTMSM) or without (QST, MGSMMSGQSTMSM) extended glycosylatable tag represented as a white lettered rectangle with a Y symbol or just as a white lettered rectangle, respectively. An extra glycosylation site was placed in the residue 105 by Q105N replacing. TMDs are colored as in Figure 66A. (B) *In vitro* translation of the 134mer truncated variants in the presence of rough microsomes. Lane 1, sample was treated with endoglycosidase H (EH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while one or two black dots indicates glycosylation of one or two glycosylation tags, respectively. (C) Schematic representation of the membrane topology of the 134mer truncate. The position of the glycosylatable tag and its glycosylation status (white and black dots) is also shown. Two potential topologies for the 134mer are represented.

Finally, translation/insertion experiments for truncated constructs 168mer (helices A to E, Figure 70A) and 205mer (helices A to F, Figure 70A) showed singly and doubly glycosylated forms, respectively (Figure 70B, lanes 3 and 6). These results suggest the efficient insertion of the helix E (Figure 70C, top). In contrast, the glycosylation pattern observed for 205mer NST (~80% of singly-glycosylated form) suggests an inefficient insertion of helix F (Figure 70C, center), as occurred with helix C. It should be noted that the low insertion efficiency was somehow expected according to its high  $\Delta G$  positive value (Figure 66B). Interestingly, translation of full-length bR construct harboring the glycosylatable N- and C-terminal tags (Figure 70A) rendered singly-glycosylated molecules (Figure 70B, lane 9), supporting the insertion of both, helices F and G into the ER membrane (Figure 70C, bottom). Apparently, the protein remains topologically ‘uncompleted’ until the last TMD has been synthesized. These results are not surprising since Kokubo & Okamoto proposed based on simulations of electrostatic interactions, that the helix G assemble with helix B with the help of the other helices to acquire the native configuration (Kokubo & Okamoto 2009).

## Results



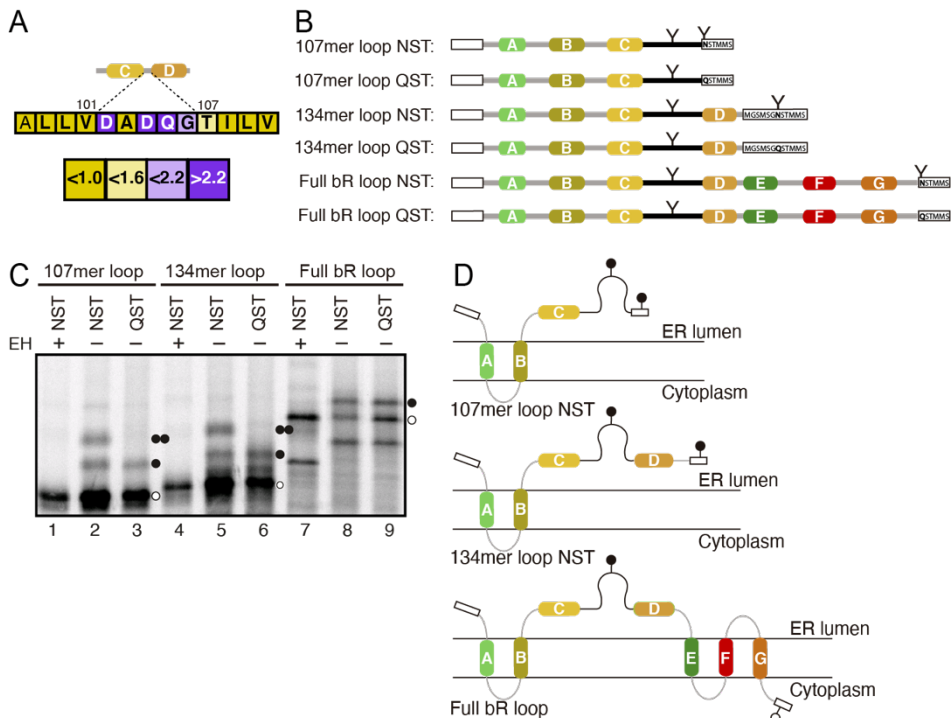
**Figure 70 | Glycosylation mapping of A-E, A-F, and full-length bR.**

(A) Schematic representation of the A to E (168mer) and A to F (205mer) truncates, and full-length bR used in the assay. To monitor the membrane orientation of truncated bR molecules a glycosylatable (NST, NSTMSM) tag (white lettered rectangle with a Y symbol) was added at position 168 (168mer), 205 (205mer), and at the end of the sequence (full bR). Simultaneously, non-acceptor (QST, QSTMM) tag (white lettered rectangle) was placed to ensure a correct topology characterization. TMDs are colored as in Figure 66A (B) *In vitro* translation of the 168mer and 205mer truncated variants, and full-length bR in the presence of rough microsomes. Lanes 1, 4, and 7, sample was treated with endoglycosidase H (EH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while a black dot indicates glycosylation of the glycosylation tag. (C) Schematic representation of the membrane topology of the 168mer and 205mer truncates, and the full-length bR. The position of the glycosylatable tag and its glycosylation status (white and black dots) is also shown. Two potential topologies for the 168mer and its abundances are represented.

Taking into account the results observed in Figure 69, we hypothesized that helices C and D insert together as a helical hairpin once it is finished the translation of helix D. Thus, the full membrane insertion of TMDs C and D is apparently dependent on the short loop connecting both TMD. To investigate the relevance of this connecting region, an analysis of the turn-inducing propensity of the residues

found in this loop was performed (Monné et al. 1999). The results revealed that four of the five amino acid residues interconnecting C and D helices are high turn inducers (Figure 71A). The presence of those amino acid residues may help to the insertion of the helices C and D regardless of their low predicted hydrophobicity (Figure 66C). To analyze the effect of this short loop between C and D in the helix insertion capabilities, this region was enlarged by inserting the N-terminal (without the SS) sequence of the berberine bridge enzyme (*bbe*) from *Papaver somniferum*, as previously done for a plant viral p9 protein (Saurí et al. 2009). This large fragment facilitated the topological analysis by virtue of its length and its glycosylation site (Figure 71B, see the sequence in Annex VI). This protein domain has been demonstrated to be translocated and concomitantly glycosylated into the microsomal lumen (Abell et al. 2002). To study the topogenesis of this bR region in greater detail, a series of truncated intermediates were generated. First, the effect of the *bbe* domain (107mer loop NST) resulted in no difference with non-extended loop 107mer (compare Figure 71C, lane 2 and Figure 68B, lane 4). Interestingly, when a bR truncated molecules that included the first three helices followed by *bbe* loop plus helix D (134mer loop) was translated an observable difference was detected (compare Figure 69B, lane 4 and Figure 71C, lane 5). Whereas the non-extended version of the 134mer was capable to insert efficiently, the loop extended version (134mer loop) did not insert properly (compare Figure 69C and Figure 71D, center). Please note that, despite the same pattern of double glycosylation in both, N-tag-134mer QST and 134mer loop NST, the second glycosylation band in the loop-extended truncate comes from the glycosylatable C-terminal tag (Figure 71C, lanes 5 and 6). Finally, when the full-length bR harboring the loop between helix C and D was translated, a singly glycosylated band was detected (Figure 71B, lanes 8 and 9), implying that the final folding is no longer acquired due to the destabilization between helix C and D (Figure 71D, bottom). These results indicate that the short loop between helices C and D is essential for bR insertion and folding into the ER membrane. Together with the hydrogen-bonds side-chain interactions between the bR helices C and D (Joh et al. 2008), the short loop found in the wild-type sequence may play role in the stabilization of the protein since the amino acid residues involved are turn inducers.

## Results



**Figure 71 | Glycosylation mapping of bR truncates harboring an extended loop connecting helices C and D.**

(A) Turn-inducing propensity at the interhelical region of bR C-D hairpin according to the scale of Monné and coworkers (Monné et al. 1999). Residues are highlighted according to their turn potential: yellow for a potential lower than 1 (Ala, Leu, Val, and Ile), light yellow for a potential between 1 and 1.6 (Thr), light purple for a potential between 1.6 and 2.2 (Gly), and purple for a turn potential above 2.2 (Asp and Gln). (B) Schematic representation of the TMD A to C (107mer loop) and A to D (134mer loop) truncates, and full-length bR loop used in the assay. All those constructs are derived from the bR full-length with the extended berberine bridge enzyme (*bbe*) loop (black) between the helix C and D. To monitor the membrane orientation of truncated bR molecules a glycosylatable (NST, NSTMSM) tag (white lettered rectangle with a Y symbol) was added at position 107 (107mer loop), 134 (134mer loop) and at the end of the sequence (full bR). Simultaneously, non-acceptor (QST, QSTMSM) tag (white lettered rectangle) was placed to ensure a correct topology characterization. TMDs are colored as in Figure 66A. (C) *In vitro* translation of the 107mer loop and 134mer loop truncated variants and full-length bR loop in the presence of rough microsomes. Lanes 1, 4, and 7, sample were treated with endoglycosidase H (EH) prior to SDS-PAGE analysis. A white dot marks the non-glycosylated form of the protein while a black dot indicates glycosylation of the glycosylation tag. (D) Schematic representation of the membrane topology of the 107mer loop and 134mer loop truncates, and the full-length bR loop. The position of the glycosylatable tag and its glycosylation status (white and black dots) is also shown.

## **DISCUSSION**



Advances over the last decades have revealed much about how the structural properties of the membrane and the protein biogenesis specialized machinery are linked to the thermodynamics of MP folding. In the present thesis part of this folding is addressed by studying the firsts stages of protein biosynthesis inside the ribosome channel, by characterizing the amino acid ‘biological’ preference for the membrane interface, by studying the effect of the hydrophobic mismatch on MP oligomerization in an *in vivo* context, and by studying how some MPs achieve their membrane topology.

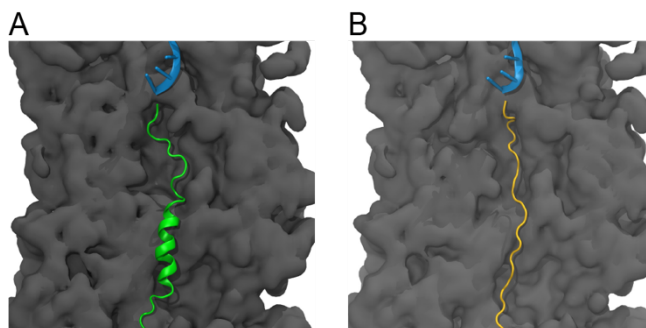
The ribosome exit tunnel (in both prokaryotes and eukaryotes) is ~100 Å in length, varying in diameter between a maximum of 20 Å distal the P-site and a minimum of a 10 Å at its narrowest constriction ~830 Å away from the P-site (Gumbart et al. 2011; Voss et al. 2006). Several previous studies demonstrated that the environment of the ribosomal tunnel is permissive for the adoption of compact (likely  $\alpha$ -helical) structures of some nascent polypeptide chains (Mingarro et al. 2000; Bhushan et al. 2010; Woolhead et al. 2004; Lu & Deutsch 2005b; L. W. Tu & Deutsch 2017) along its entire length, including at the constriction (T. Su et al. 2017) and very close to the ribosomal P-site (Lin et al. 2011). It has also been shown that limited tertiary folding can take place, but possibly only in the wide exit domain region of the tunnel distal to the P-site (Hedman et al. 2015; Fariás-Rico et al. 2017; Kosolapov & Deutsch 2009; L. Tu et al. 2014). However, not all sequences that adopt an  $\alpha$ -helical structure in the folded protein necessarily would form within the ribosome.

In the present thesis, we have studied the intrinsic ability of helical sequences in the context of stalled nascent polypeptide chains to attain a compact structure within the ribosome. Peptide sequences with known stable helical propensity in their native folds from both integral membrane proteins (hydrophobic helices) and soluble proteins (hydrophilic helices) were used. A principal finding was that when hydrophobic test sequences were present within the ribosomal tunnel there was clear evidence of compaction, as demonstrated by reduced glycosylation efficiency at the ‘critical’ distance, whereas there was no such indication for hydrophilic test sequences despite their known  $\alpha$ -helical propensity. All of the six hydrophobic tested sequences (VSV-G, gp41, GpA, LepH1, p75, and M13) showed compaction, indicating that nascent chain  $\alpha$ -helix accommodation in the ribosome is a common, if not a general phenomenon. This is in agreement with earlier studies using FRET to measure compaction of hydrophobic and hydrophilic sequences showing that hydrophobic sequences compacted whereas hydrophilic sequences did not (Woolhead et al. 2004). However, the hydrophilic sequences used in this earlier study were not specifically selected on the basis of helical propensity, raising the possibility that lack of compaction was simply a property of a non-helical sequence being tested. The positioning of the tested sequences in the FRET study was closer to the P-site

## Discussion

(four residues away) than in the present thesis in which the start of the test sequences is located 7-9 residues from the P-site (for nascent chains with  $d=67$  residues). This suggests that the adoption of  $\alpha$ -helical conformation for hydrophobic helices is a very early event initiated 20-30 Å away from the P-site.

What is it about the environment within the ribosome that induces hydrophobic, but not hydrophilic, helices to adopt helical conformation? It has been previously hypothesized that non-polar areas within the tunnel may provide a surface for hydrophobic helices to nucleate upon (Liao et al. 1997). The cryo-EM structure does not reveal any obvious extensive hydrophobic surface for this to happen (Nissen et al. 2000). MD simulations of the test peptides in the ribosome were performed to try to address the above question. The main difference observed between tested sequences was a reduction in solvent accessible surface area of hydrophobic sequences in comparison to hydrophilic sequences when modeled as helices. The simulations suggest that hydrophobic amino acids within ribosomal proteins provide hydrophobic contacts with amino acid residues in hydrophobic helices (Figure 72). Alternatively, it has been shown theoretically that space confinement within a cylinder resampling the ribosomal tunnel can entropically stabilize  $\alpha$ -helices without invoking-specific interaction with the tunnel wall (Ziv et al. 2005).



**Figure 72 | Nascent polypeptide conformations within the ribosome exit tunnel.**

Screenshot of molecular dynamic simulations of helical sequences of folded gp41 (green, A) and unfolded NAGK (orange, B) within the ribosome exit tunnel (gray) shown in cartoon representation. Solvent accessible surface area (SASA) was calculated from those simulations. tRNA bound to the P-site is shown in blue.

In the experiments where two Ile residues in the mid-region of VSV-G were replaced with amino acid residues intended to disrupt helicity or reduce overall hydrophobicity, there was in general a loss of compaction. There were, however, quantitative differences with Pro-Pro > Lys-Lys > Gly-Gly > Asp-Asp > Ala-Ala with respect to increasing glycosylation efficiency, a measure of loss of helicity. Furthermore, an incomplete VSV-G TMD did not compact (Figure 33), in good agreement with previous FRET measurements (Woolhead et al. 2004). Although the



precise number of residues in helical conformation is unknown, experiments estimate that at least 14 hydrophobic residues from the VSV-G TMD sequence appear to be needed to display a glycosylation pattern compatible with  $\alpha$ -helix formation, as well as for membrane integration (Figure 33). These results suggest that a long hydrophobic stretch of amino acids is a requirement to facilitate stable helix formation in the ribosome. This does not rule out localized transient short helix formation but indicated that a long helix can make more substantial contacts and/or increase entropic stabilization, sufficient to stabilize the helical conformation (Ziv et al. 2005). There must be some cooperativity of interactions within the helix, or with the walls of the ribosome tunnel that facilitate the stabilization. Is there a physiological relevance for hydrophobic helices forming within the ribosomal tunnel? For spontaneous insertion of integral membrane proteins into membranes, it has been proposed that proteins must adopt helical conformation pre-insertion to overcome unfavorable energetic barrier if exposing polar peptide bonds to the hydrophobic interior of a lipid bilayer (Popot & Engelman 2000; White & G. von Heijne 2008). It is possible that this is also the case for translocon-assisted insertion and that folding in the ribosome enhances recognition of TMD helices for integration. Without the early adoption of a helical conformation, TMDs would enter the translocon with exposed peptide bonds which may negate recognition for integration. So, it can be remarked that overall hydrophobicity, helicity, and length are major determinants of  $\alpha$ -helical adoption within the ribosomal tunnel. This could facilitate recognition by SRP (Voorhees & Hegde 2015) and/or a favorable conformation for membrane integration upon entering the translocon.

As we just discussed, sufficiently hydrophobic regions are folded in the ribosome exit tunnel and will be able to partition into the membrane. On the other hand, less hydrophobic sequences that are unable to insert into the membrane due to thermodynamic restrictions may take a path towards the membrane's interfaces. We have to take into account that small hydrophobic peptides that cannot form secondary structures must be restricted to the interface because of the energetic cost of partitioning the peptide bonds into the hydrocarbon core (Jayasinghe et al. 2001).

The biological membranes are not formed exclusively by a highly hydrophobic hydrocarbon region, they also present interfacial regions. Of the total thickness of the bilayer, the two interfaces account for approximately 50% (~30 Å) of total membrane thickness (Figure 3D). This emphasizes the fact that the bilayer should not be simply conceived as a thin hydrocarbon slab separating two aqueous phases. The presence of a mixture of water and polar head groups generates non-covalent interactions between the interfaces and peptides (White & Wimley 1998). The 15 Å thermal thickness of an interface can comfortably accommodate an  $\alpha$ -helix (diameter ~10 Å) running parallel to the bilayer plane. Despite the importance of the membrane

## Discussion

interface, the study of how peptides are influenced to match this region is currently uncompleted.

During this thesis, we developed a novel tool to determine the surface disposition of short peptide sequences. To setup the approach, the MGD for interfacially placed sequences was firstly determined. The sharp threshold for *N*-linked glycosylation from the ends of an interfacially located sequence and the OST catalytic site turned to be 8-9 residues from the N-terminus of the sequence and 10-11 residues from the C-terminus end. These distances are shorter than those estimated for membrane-inserted sequences (Nilsson & G. von Heijne 1993), probably due to the closer proximity to the OST catalytic site to the membrane interface rather than the hydrophobic core.

Additionally, we developed a Lep3G variation that includes 3 different glycosylation acceptor sites (G1, G2, and G3) in the same construct, providing a better tool to measure interfacial disposition, as the results with pHLIP3D and melittin derived sequences showed.

The Lep3G variation also provided us the possibility to determine the interfacial propensity of the 20 naturally occurring amino acids, since the fraction of singly, doubly, and triply glycosylated Lep3G-derived molecules can be expressed as an apparent equilibrium constant ( $K_{app}$ ) between surface, inserted and translocated states. Each amino acid displayed a preference for one of the three possible states, as summarized in Figure 45. Here, we provide a complete interfacial scale based on the amino acid preference for the three different states in an interfacial background. A direct comparison between the Lep3G-derived apparent free energy values with a previous interfacial propensity scale (Wimley & White 1996) showed a similar distribution (Figure 43). Nevertheless, not a perfect correlation should be expected due to methodological differences between both approaches. Wimley and White calculated the free energies to transfer peptides from water to POPC interfaces, while our approach uses biological membranes from rough microsomes. In any case, both scales share similarities being the Trp the most interfacial residue, while charged residues are the less interfacial prone. Ala residue also showed an interfacial preference, in good agreement with previous data that determine its preference for the lipid-water interface region over the central acyl chain part of the ER membrane (White & Wimley 1999; Nilsson et al. 2003). Our results also match with structure-based statistical analysis of amino acid distribution in TMD helices (Baeza-Delgado et al. 2013) (Figure 22). As an example of the similarities, aromatic residues are usually found at the ends of the TMDs. These regions are placed in the membrane interface, a space created by the headgroups between the hydrophobic core and the bulk water.

Despite amino acid propensity to one of the three states, it should be mentioned that the presence of it in the other two states is also possible. The calculated apparent free energy values only reflected the intrinsic preference of any amino acid by the state, not being the unique “option” unless the  $\Delta G_{app}$  show a high absolute value. As an example, Tyr, a residue with a slight preference for the inserted state versus surface state in our data (Figure 41), could be found at the end of the TMD, like other aromatic residues like Trp, but also more centrally located (Figure 22), denoting the aforementioned slight preference for inserted state.

With regard to charged residues, all displayed preference for translocated state. However, further examination of the apparent free energies between the surface and inserted states  $\Delta\Delta G_{app}^{S/I}$  (Figure 41B) reveals differences between both, positive and negatively charged residues. Positively charged residues (Arg and Lys) displayed a certain tendency to the surface state, whereas negatively charged residues (Asp and Glu) preferred the inserted state versus the surface state. This effect could be influenced by the Charge Balance Rule (Dowhan et al. 2019), i.e., the negatively charged lipid headgroups could benefit the interfacial disposition of positively charged residues. Hence, positively charged residues will be prone to fit in the membrane interface whereas, negatively charged residues (likely after protonation) prefer to be inserted rather than in the membrane interface due to the electrostatic repulsions between similar charges. Note that His residue is probably less affected by the charge effect because its  $pK_R$  is around 6.0 and at physiological pH 7.4 the side chain remains partially uncharged.

The determination of the amino acid interfacial propensity not only allows peptide screenings but also helps to *de novo* design of sequences with amphipathic properties. An application of this principle was the replacement of the Bax’s non-conserved amino acid S118 (Xiao et al. 2016) since Ser is a residue with low interfacial propensity (Figure 44). Through these changes, the rationally modified BaxE5 peptide was capable of disturbing the integrity of lipid vesicles in a pH-dependent manner.

As we are discussing, amino acid composition and its preferences for different membrane regions influence the folding of proteins but, what happens next? By inserting into the membrane, TMDs necessarily create a local environment less hydrophobic than the surrounding bulk lipid (Engelman et al. 2003). This less favorable local environment may be exacerbated by hydrophobic mismatch between TMD length and bilayer thickness. In the complex scenario of the membrane, the exposure of hydrophobic residues and lipid groups to water is highly unfavorable and therefore expected to be reduced. Helices in TMDs are, on average, 24 amino acids long (36 Å), ranging from 17 to 34 amino acids long (25.5-51 Å) (Baeza-Delgado et al. 2013). A stretch of approximately 20 consecutive hydrophobic amino acids can

## Discussion

span 30 Å of the hydrocarbon core of a model biological membrane. Indeed, the most prevalent (~12%) length in TMDs is 21 residues (Baeza-Delgado et al. 2013). Based on previous *in vitro* work, it was proposed that longer helices can span the bilayer with a concomitant tilting of the helix axis relative to the membrane plane (Strandberg et al. 2012). Other options to accommodate the wide variety of TMDs range from lipid accommodation to helix deformation. Alternatively, each TMD can be located in a lipid bilayer or a lipid bilayer sub-domain that matches its hydrophobic length, thereby minimizing peptide adaptations. In fact, the subcellular distribution of helical membrane proteins among multiple organelles based on their hydrophobic lengths is considered to occur (Cosson et al. 2013).

GpA offers a valuable tool to study the influence of hydrophobic matching on helix-helix packing in biological membranes, even in living cells. Not only successful formation of a GpA dimer depends on the correct disposition of the TMD, but it has also been demonstrated *in vitro* that the hydrophobic mismatch influences TMD packing in micelles (Orzáez et al. 2000; Orzáez et al. 2005). Furthermore, the NMR structure of the homo-dimer was solved both in detergent micelles (MacKenzie et al. 1997) and in membrane bilayers (Smith et al. 2001) and the dimerization motif has been thoroughly studied *in vitro* (Lemmon et al. 1992). Given the aforementioned characteristics of GpA, it was used to challenge hydrophobic matching under *in vivo* conditions, being the first study in which hydrophobic mismatch has been systematically explored in biological membranes.

All analyzed TMDs, ranging from 17 to 29 hydrophobic residues in length (25.5-43.5 Å), including a minimized GVxxGVxxT dimerization motif were capable of homo-oligomerize *in vivo*, both in prokaryotic and eukaryotic membranes. It is worth mentioning that based on our studies, each of these TMDs could oligomerize in membranes from different organelles (in eukaryotic cells) or membrane domains with distinct lipid composition and consequently different hydrophobic core length. In the ToxRED assay, the presence of the GpA minimal dimerization motif turned to be not absolutely necessary for the oligomerization of long TMDs. It seems that in *E. coli* membranes the positive, but not the negative, hydrophobic mismatch drives the association of TMDs. In human-derived cells, a similar scenario could be observed. However, in this case, Gly to Ile replacement in long TMDs does have an impact on their dimerization (Figure 51C). These results may suggest that eukaryotic membranes are more adaptable to a hydrophobic mismatch. While in HEK-293T cells the membrane can adapt to a positive mismatch scenario, in *E. coli* the TMD is mainly responsible for the elimination/reduction of the free energy associated with the exposure of hydrophobic residues to the aqueous environment.

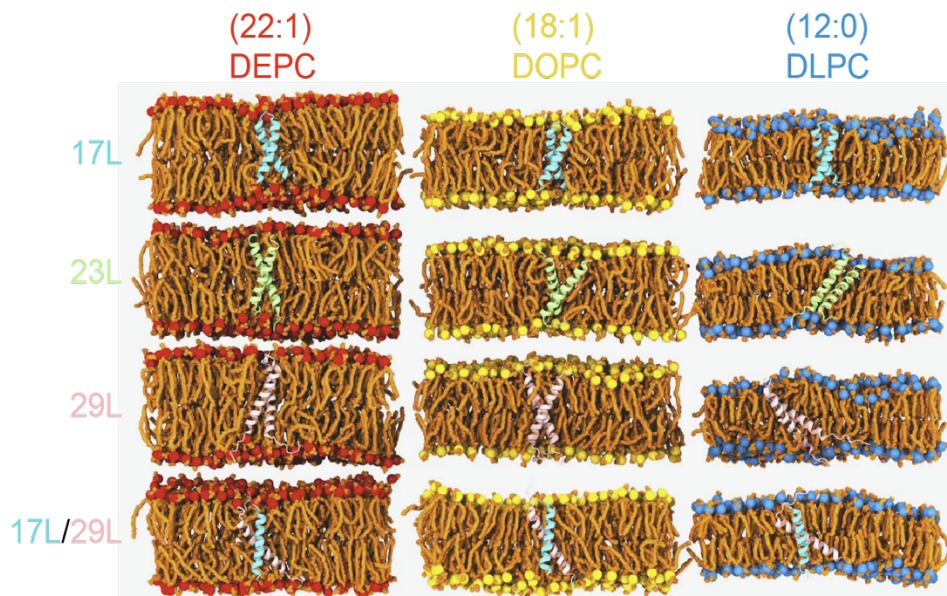
When hetero-dimer formation was assayed (in eukaryotic cells using BiFC assay) between chimeras with different hydrophobic length, it was found that most

of the analyzed combinations returned fluorescence values significantly above the controls (Figure 52). These results indicated that cell membranes are capable of hosting a dimer between two TMDs with large hydrophobic length disparity among them. Nonetheless, significant differences in the intensity of the studied heterodimers were observed. Once again, these variations could have a multi-factorial origin. Either the length of the hydrophobic segments influences partially the organelle or the membrane sub-domain localization, or the disparity in the hydrophobic length has a penalty on dimer formation, which would, for the first time, show an effect of the hydrophobic match between the TMDs and the membrane on the packing of TMD in living cells. Based on the localization assays, the last option seems the most plausible explanation. When the subcellular localization of the chimeras was studied it was found that all of them colocalized with ER marker (Figure 57) but not with a PM marker (Figure 58), suggesting that ER membranes can adjust to the hydrophobic mismatch better than the PM. This unexpected and homogeneous localization could also take place because dimerization precludes protein sorting, preventing the newly formed chimeras escaping from the ER. Nonetheless, the full-length GpA micrographs (Figure 59) indicate that the wild-type sequence can sort from the ER in HEK-293T cells despite forming dimers. Full-length GpA sorting can probably be achieved through its native signal sequence, which was not included in the chimeras (Figure 57 and Figure 58).

The atomistic MD simulation data performed by the collaborators indicated that 17L, 23L, and 29L homo-dimers are stable regardless of the hydrophobic thickness of the bilayer, while the dimerization motif deficient polyL control dissociates rapidly. However, the 23L and 29L peptides required some adaptation, especially in DLPC membranes where the mismatch is large and positive (Figure 73). The residues near the interface showed less helicity and more flexibility while keeping the structure of the dimer motif intact. Furthermore, the 23L and 29L homo-dimers in a DLPC membrane were characterized by large tilt angles. All these changes took place to keep the TMD within the limited hydrophobic thickness of the lipid bilayers. However, no elongation of the dimer in terms of a structural change from a canonical  $\alpha$ -helix to the  $3_{10}$  helix structure was observed. Yet, a perturbation of membrane thickness near the dimer was detected. For the 17L dimer, this perturbation increased upon increasing negative mismatch (DEPC>DOPC>DLPC). For the 23L and 29L dimers, the effect was stronger or comparable for DOPC than DEPC membranes. Interestingly, in the thinner DLPC bilayer, little lipid adaptation was observed since the entire dimers tilted substantially to match the hydrophobic thickness. This suggests that there is a maximum for lipid adaptation, beyond which tilting and peptide deformation become the preferred mechanism for adaptation. The behavior of the 17L/29L hetero-dimer as a whole was similar to a 23L homo-dimer, while the structures of the individual peptides resembled those in the corresponding homo-dimers. Importantly, the simulations are able to probe degrees of mismatch larger

## Discussion

than those appearing in the *in vivo* experiments. In the DOPC membrane, the thickness of which likely matches that of the *E. coli* and HEK-293T membranes studied *in vivo*, the differences observed in dimer tilting were actually small. Therefore, the importance of this tilting in living membranes might not play a major role, and the perturbation of membrane thickness is likely the dominant adaptation mechanism *in vivo*.



**Figure 73 | Molecular dynamic simulations of homo and hetero-dimers.**

Screenshot of molecular dynamic simulations of homo and hetero-dimers embedded in model membranes with different lipid composition (DEPC, 1,2-dierucoyl-sn-glycero-3-phosphocholine (red headgroups), DOPC, 1,2-dioleoyl-sn-glycero-3-phosphocholine (yellow headgroups), and DLPC, 1,2-dilauroyl-sn-glycero-3-phosphocholine (blue headgroups)) to provide differences in hydrophobic thickness of the membrane core (31.6 Å, 38.7 Å, and 47.3 Å, respectively). Different adaptations are present according to the hydrophobic thickness of the membrane and the TMD length (17 residues (light blue), 23 residues (green), and 29 residues (pink)).

The stability of the native GpA dimers in various lipid environments has been studied in previous works using both coarse-grained (Flinner et al. 2014; Sengupta & Marrink 2010; Janosi et al. 2010) and atomistic (Hénin et al. 2005) simulations. These atomistic simulations reported small differences in the dimeric structure of native GpA embedded in membranes of different thicknesses. The simulation data complemented this information from a novel perspective, now discussing the variation of the peptide lengths as a new variable, therefore covering larger values of mismatch and by reducing the importance of the structural features of the GpA dimer to the minimized dimerization motif. Despite these extensions, our results agree with the earlier work on GpA (Petrache et al. 2000).

The two adaptation mechanisms resolved here have been previously reported in systematic computational studies on KALP and WALP monomers of different lengths embedded in membranes of various thicknesses (T. Kim & Im 2010; Kandasamy & Larson 2006). It seems that in the case of stable dimers such as those studied here, the adaptation mechanisms are similar to the ones observed previously for peptide monomers. It must be noted that some other suggested mechanisms (Kandasamy & Larson 2006), such as the formation of a non-lamellar phase or the expulsion of the peptides from the membrane, are not feasible in our simulation study and achievable timescales. Furthermore, our atomistic study is limited to individual dimers. Therefore, we cannot rule out the possibility that membrane stress is relieved through the formation of larger oligomers. The present thesis not only highlights the capability of biological membranes to host TMD homo-oligomers with hydrophobic regions ranging from 25.5 to 43.5 Å, but it also demonstrates the ability of lipids and peptide regions to cooperate in order to minimize the hydrophobic mismatch.

All the previous information and techniques can also contribute to understanding the final topology acquisition of MPs. Hence, we analyzed the biogenesis of TRAP $\gamma$  subunit and bR taking advantage of the glycosylation mapping assay. The TRAP $\gamma$  results demonstrate that this subunit of the TRAP complex consists in four TMDs with its N- and C-termini facing on the cytosolic side, which is consistent with the *in silico* predictions but reverse to the topology suggested in a seminal work (Hartmann et al. 1993). It is also noticeable that the subunit contains a poorly hydrophobic TMD (TMD4) whose insertion capacity notably increases in the presence of its preceding TMD1 and/or TMD2. This effect could be due to a tertiary structure interaction or a temporary cooperation between these hydrophobic regions at the time of their insertion (Sadlish et al. 2005; Saurí et al. 2007) that cannot be concluded from our results. In the case of bR, our results corroborate the well-known seven TMDs with an extracellular N-terminus and a cytoplasmatic C-terminus (Kahn & Engelman 1992). As in TRAP $\gamma$ , some of the TMDs are poorly hydrophobic (helix C, D, F, and G) and their insertion capabilities uplift in the presence of the adjacent TMD. In the case of helices F and G, when helix G finally arrives to the membrane, the final topology of bR is acquired (Figure 70B). The presence of helix D is also a requirement for the correct insertion of helix C (Figure 69C). Nevertheless, not only the presence of helix D but also the presence of the short connecting loop between these helices is a requirement for the proper insertion of these two helices as a helical hairpin (Figure 71B). A loop extension probably does not allow a spatial contact during the insertion process, disfavoring the helix packing and thus, precluding insertion. Taking all together the MP biogenesis and final topology acquisition require the presence of all TMDs and, likely, some additional cellular components as chaperones that provide an adequate environment to get the protein native conformation at least for the two proteins studied in this thesis.





# CONCLUSIONS



1. Hydrophobicity, helicity, and length are major determinants for  $\alpha$ -helical conformation adoption within the ribosome exit tunnel.
2. The development of the Lep3G assay provides a straightforward method to determinate interfacial disposition in a systematic and quantitative manner. Therefore, we have determined an interfacial ‘biological’ scale for the 20 naturally occurring amino acids.
3. Biological membranes host homo- and hetero-oligomers with largely different hydrophobic regions, emphasizing the ability of lipids and proteins to cooperate in order to minimize the hydrophobic mismatch.
4. Insertion and final membrane protein topology are not only guided by the hydrophobicity of the TMD, but also by TMD-TMD packing and the presence of specialized connecting loops.



## **RESUMEN**



## Introducción

Entre las principales moléculas necesarias para realizar las funciones celulares encontramos a las proteínas. Las proteínas presentan una gran variedad de grupos funcionales y estructurales, entre los que podemos encontrar el grupo de las proteínas de membrana (MPs). Las MPs son proteínas que están asociadas y/o ancladas a las membranas biológicas y están ampliamente distribuidas en todos los seres vivos (aproximadamente el 30% de los genes humanos codifican MPs (Lander et al. 2001; Venter et al. 2001)). La principal característica que presentan las MPs son sus dominios transmembrana (TMDs), capaces de plegarse, insertarse y funcionar en el entorno hidrofóbico de la membrana. En base a la estructura secundaria de estos TMD, podemos dividir las MPs en dos grandes grupos, barriles  $\beta$  y hélices  $\alpha$ .

A pesar de su abundancia y necesidad para la vida (sin ellas las membranas biológicas simplemente serían barreras lipídicas herméticas), el conocimiento que se tiene sobre las MPs está lejos del que se tiene de proteínas solubles. Es por ello por lo que la presente tesis va a centrar su atención en el estudio de las MPs, más específicamente en el grupo de las MPs cuya estructura se basa en hélices  $\alpha$ .

Durante su biogénesis, tanto proteínas integrales de membrana como de secreción deben atravesar la membrana parcial o totalmente. Para que esto suceda se requiere de una maquinaria especializada que asiste este proceso, evitando el plegamiento de las proteínas hasta que su inserción o translocación de la membrana ha concluido. Este proceso puede llevarse a cabo tanto cotraduccionalmente (mientras se sintetizan en los ribosomas) como postraduccionalmente (una vez se han separado la cadena polipeptídica del ribosoma), dependiendo de la características físico-químicas de las proteínas.

El ribosoma, maquinaria celular encargada de la síntesis proteica, está compuesto por proteínas y RNA que se organizan en dos subunidades de diferente tamaño. La subunidad pequeña se encarga de la fidelidad del proceso mediando la interacción entre los anticodones del tRNA (RNA de transferencia) y los codones del mRNA (RNA mensajero). La subunidad grande se encarga, entre otras cosas, de la formación del enlace peptídico en el sitio P del ribosoma. Una vez formado el enlace peptídico, la cadena naciente es liberada al exterior del ribosoma a través de un conducto denominado túnel de salida del ribosoma. Debido a las características de este túnel (una longitud de 80-100 Å, un grosor de  $\sim 28$  Å) proporciona protección a las cadenas nacientes. Dadas las dimensiones del túnel, se ha propuesto que tiene una participación activa en el plegamiento de estructuras secundarias como hélices  $\alpha$  o incluso estructuras terciarias como pequeños dominios denominados dedos de zinc (Mingarro et al. 2000; Woolhead et al. 2004; Bhushan et al. 2010; Hedman et al. 2015).

## Resumen

Una vez las cadenas nacientes salen del túnel del ribosoma su direccionamiento hacia la membrana puede ocurrir cotraduccional o postraduccionalmente. La ruta más común y ampliamente estudiada es la vía cotraduccional. Para esta vía se requiere la participación de la partícula de reconocimiento de la señal (SRP), encargada de capturar la cadena naciente según emerge del ribosoma y llevarla a la membrana del retículo. Es en este punto donde el ribosoma, todavía en el proceso de traducción, interacciona con el canal a través del cual se realiza la inserción o translocación, el denominado translocón.

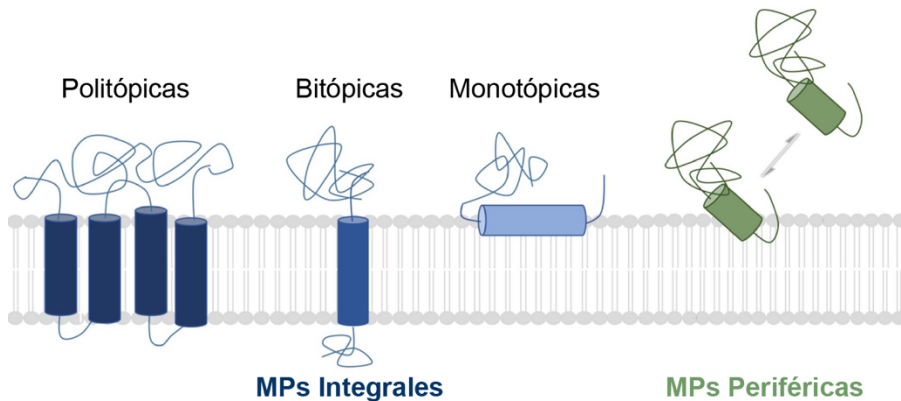
El principal componente del translocón, en eucariotas, es el complejo Sec61. Este complejo multiprotéico forma el poro por el cual las proteínas son insertadas en la membrana o translocadas al lumen del retículo. La subunidad Sec61 $\alpha$  es el principal componente del complejo, formando el canal propiamente dicho. Este canal es el único encontrado en las membranas biológicas que permite el paso de moléculas en dos direcciones, de un lado al otro de la membrana y lateralmente. Para ello, presenta un poro central y una apertura lateral por la cual los TMD quedan expuestos directamente a la membrana lipídica, facilitando su inserción. Adicionalmente, el translocón tiene asociadas otras proteínas accesorias entre las que se destaca la oligosacariltransferasa (OST). La OST es una glicosiltransferasa encargada de transferir un tetradecasacárido a un residuo de Asn en una secuencia de Asn-Xaa-Ser o Asn-Xaa-Thr, donde Xaa es cualquier amino ácido excepto Pro.

Pese a que casi la totalidad de las MPs utilizan la maquinaria especializada del translocón para insertarse, la topología que adquiere cada una depende de su propia secuencia y de los determinantes topológicos que presente. Globalmente podemos clasificar las proteínas integrales de membrana en dos tipos básicos, las que atraviesan una única vez la membrana (bitópicas) y las que la atraviesan en múltiples ocasiones (politópicas) (Figura 1). Las proteínas que atraviesan una vez la membrana pueden adoptar diferentes topologías según la presencia o ausencia de péptido señal y en lo que a la orientación de la proteína madura se refiere. De igual forma esto sucede con las proteínas que poseen varios TMDs y atraviesan en múltiples ocasiones la membrana, siendo clave para la topología de éstas la disposición que adopte el primer TMD en su secuencia.

Además de estos 2 tipos, existen proteínas que se asocian a la membrana sin atravesarla, las denominadas MPs monotópicas o MPs interfaciales, se encuentran asociadas a una única cara de la bicapa lipídica (Figura 1). Los mecanismos por los cuales este tipo de proteínas son direccionadas e interaccionan con la membrana no están completamente descritos debido a la dificultad de estudio en el entorno donde se encuentran. Sin embargo, el estudio de las pocas estructuras que se han conseguido resolver de proteínas interfaciales revela que una de las características más comunes



es su interacción con las membranas a través de hélices anfipáticas (A. L. Lomize et al. 2007; Pataki et al. 2018).



**Figura 1 | Representación de la estructura de las proteínas de membrana**

Las proteínas integrales de membrana (MPs integrales) se clasifican en politópicas (atraviesan la membrana en múltiples ocasiones), bitópicas (atraviesan la membrana una única vez) o monotópicas (no atraviesan la membrana, sino que se quedan ancladas a una de las dos monocapas de la membrana). Las proteínas periféricas de membrana (MPs periféricas) se asocian de forma reversible a la bicapa lipídica (parte derecha del panel). Adaptado de (K. N. Allen et al. 2019).

En cuanto a la inserción en la membrana de los TMDs, el translocón no solo actúa como facilitador, sino que es un componente esencial en el proceso de selección e inserción de TMDs. Para que el translocón posibilite la inserción de TMDs se requieren algunas características en las cadenas nascentes de polipéptidos:

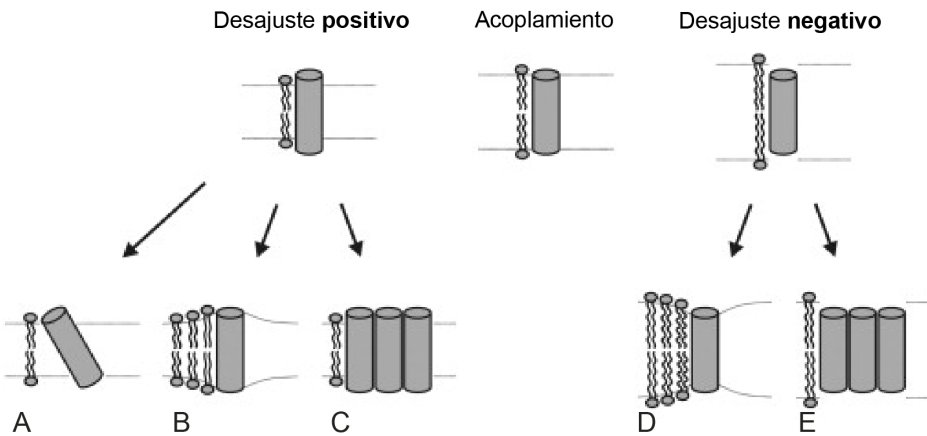
- Se requiere de un determinado grado de hidrofobicidad y helicidad en las regiones a insertar.
- La orientación de los TMDs y la presencia de determinantes topológicos.
- El grado de ajuste hidrofóbico entre el grosor de la bicapa lipídica y la longitud del TMD

El grado de hidrofobicidad se entiende como la medida de la tendencia relativa de una molécula hacia un entorno no acuoso con respecto a uno acuoso. Generalmente la hidrofobicidad de un TMD viene determinada por la cadena lateral de los residuos que componen su secuencia. La hidrofobicidad de los aminoácidos ha sido determinada experimentalmente en varias escalas, siendo la base para la detección de TMDs mediante el análisis de secuencias (MacCallum & Tieleman 2011). Entre las escalas determinadas cabe destacar por su relevancia y relación directa con la presente Tesis el estudio sistemático de inserción mediada por el

## Resumen

translocon (Hessa et al. 2005; Hessa et al. 2007), donde los autores propusieron una escala de hidrofobicidad “biológica”.

El concepto de ajuste hidrofóbico hace referencia al acoplamiento que existe entre el grosor de la bicapa lipídica y la longitud del TMD. Un desajuste hidrofóbico se produce cuando, o bien la membrana no es suficientemente gruesa para acomodar al TMD (desajuste positivo), o bien cuando la región hidrofóbica del TMD no es suficiente para atravesar el núcleo hidrofóbico de la membrana (desajuste negativo). En ambos casos el la penalización termodinámica de exponer residuos polares a la membrana (desajuste negativo) o bien de que queden expuestos residuos hidrofóbicos al ambiente acuoso (desajuste positivo) provoca que tanto los lípidos como los TMD se adapten para reducir el desajuste (Figura 2).



**Figura 2 | Posibles adaptaciones al desajuste hidrofóbico.**

En el caso de desajuste positivo: el TMD puede inclinarse para reducir la exposición de grupos hidrofóbicos al exterior de la membrana (A), los ácidos grasos pueden ordenarse incrementando el grosor hidrofóbico local adyacente al TMD (B), o los TMD pueden oligomerizar, reduciendo la exposición de residuos hidrofóbicos al ambiente acuoso (C). En el caso del desajuste negativo: los ácidos grasos pueden ordenarse, reduciendo el grosor local de la membrana (D), o los péptidos pueden oligomerizar reduciendo la exposición de residuos polares a un entorno hidrofóbico (E).

## Objetivos

El objetivo general de la tesis es estudiar la biogénesis y el plegamiento de las proteínas de membrana en el contexto biológico del translocón. Durante la tesis se han abordado los siguientes objetivos concretos:

- Caracterización de la formación de estructuras helicoidales en el túnel del ribosoma.
- Desarrollar una aproximación molecular que permita determinar de forma cuantitativa la disposición interfacial en la membrana de segmentos específicos de proteínas.
- Establecer una escala de interfacialidad de los 20 aminoácidos naturales.
- Estudiar el efecto del acoplamiento hidrofóbico en el empaquetamiento entre TMDs en membranas eucariotas.
- Determinación de la topología de la subunidad gamma del complejo TRAP.
- Estudio comparativo de la topogenesis de la proteína bacteriorodopsina con su estructura en la membrana.

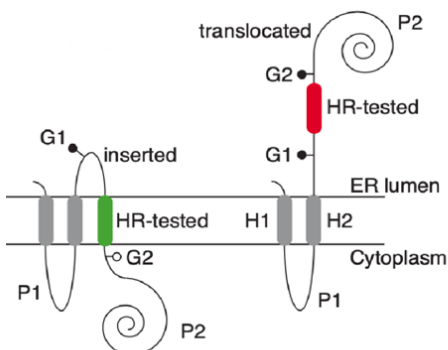
## Resumen

## Metodología

### *Determinación experimental de la inserción: Ensayo ‘Lep’*

Para determinar si un segmento hidrofóbico se inserta en la membrana de forma asistida por el translocón, podemos emplear el segmento de interés de forma aislada. La secuencia estudiada debe introducirse en una proteína modelo que permita el estudio de la inserción. En esta tesis se ha empleado como vehículo la peptidasa del péptido señal *Leader peptidase I* (código Uniprot P00803) de *E. coli*. Esta proteína está compuesta por 324 aminoácidos y contiene dos TMDs (H1 y H2). Los primeros 25 residuos contienen el dominio H1 seguidos de aminoácidos cargados positivamente, lo que es suficiente para insertar el polipéptido en la membrana con orientación N-terminal hacia el espacio extracelular ( $N_{out}$ ). Debido a ello, al expresarse *in vitro* en presencia de microsomas de páncreas, los dos TMDs se insertan eficientemente, quedando tanto el N- como el C-terminal orientados hacia el lumen microsomal (Nilsson & G. von Heijne 1993). El segmento hidrofóbico que va a ser testado se inserta entonces en el dominio C-terminal (P2) flanqueado por dos sitios aceptores de glicosilación (G1 y G2) (Figura 3). Si el segmento testado se transloca a través de la membrana, el dominio P2 completo quedará expuesto en el lado luminal de la membrana y los dos sitios de glicosilación serán modificados. De forma contraria, si el segmento testado es capaz de insertarse en la membrana, solo el sitio aceptor de glicosilación G1 podrá ser modificado.

La modificación de los sitios aceptores con la molécula de azúcar hace que la masa molecular de las quimeras ensayadas aumente aproximadamente 2kDa por molécula unida. De esta forma, mediante un análisis electroforético (SDS-PAGE), se posibilita la determinación precisa y cuantificable de las formas no glicosilada, mono-glicosilada y doble-glicosilada de las quimeras ensayadas.



**Figura 3 | Proteína modelo ‘Lep’.**

Los rectángulos grises (H1 y H2) se corresponden a las hélices nativas de Lep. Los rectángulos coloreados (verde o rojo en función de insertado o translocado) corresponden a las regiones hidrofóbicas ensayadas (HR tested). Las posiciones G1 y G2 se corresponden a los sitios aceptores de glicosilación. Los puntos negros indican la presencia de azúcares si están orientados al lado luminal de la membrana (ER lumen) y el punto blanco una diana no modificada.

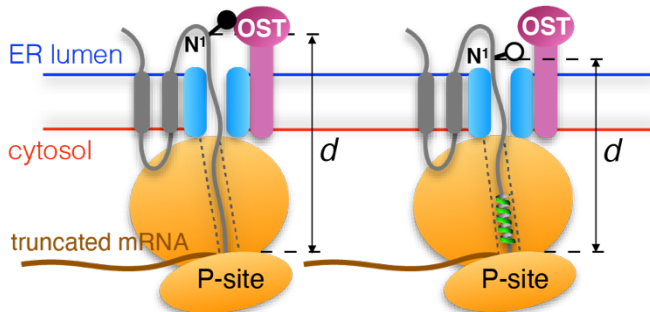
## Resumen

### *Determinación experimental del plegamiento en el interior del túnel del ribosoma*

Esta metodología es una variación del ensayo ‘Lep’. Los segmentos objeto de estudio se encuentran igualmente insertados en dominio P2 de Lep. Para determinar el plegamiento de las cadenas nascentes la principal diferencia del ensayo es la utilización de mRNAs truncados (sin codón de parada). Además, metodológicamente se ha diseñado un único sitio aceptor de glicosilación (G1) precediendo el segmento de estudio.

El ensayo se basa en la distancia mínima que requiere la enzima OST para glicosilar el sitio aceptor de glicosilación presente en dichas cadenas. Para mantener fija la distancia, las cadenas nascentes deben quedar atrapadas en el complejo formado por el ribosoma y el translocón durante la inserción cotraduccional. Para conseguir esto último los mRNAs que codifican para las cadenas nascentes carecen de codón de parada, por lo que el último residuo incorporado en el polipéptido se quedará retenido en sitio P del ribosoma. De esta forma, el número mínimo de residuos requerido para abarcar la distancia ( $d$ ) entre el sitio P del ribosoma y el centro catalítico de la OST dependerá de el grado de compactación de la porción del polipéptido situada en el interior del túnel del ribosoma (Figura 4).

Para determinar el plegamiento se realizan mRNAs truncados de distinto tamaño variando la  $d$ . Los fragmentos ensayados que se glicosilan a una  $d$  mayor son aquellos que se compactan en el interior del túnel, mientras que las secuencias que preenten una conformación extendida serán glicosiladas a valores de  $d$  menores.

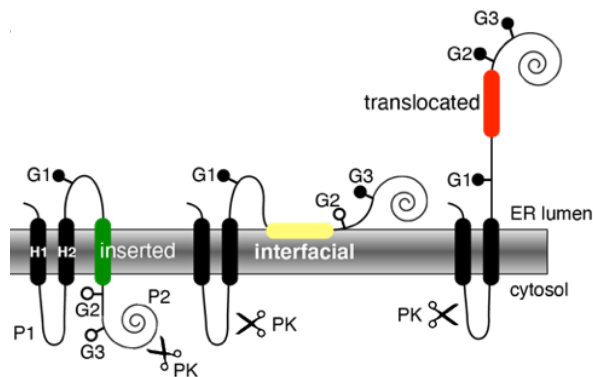


**Figura 4 | Representación del ensayo utilizado para determinar el plegamiento en el interior del túnel del ribosoma.**

La proteína modelo ‘Lep’ contiene dos TMD (rectángulos grises) y un largo dominio C-terminal. Las cadenas nascentes truncadas de diferentes tamaños de longitud unidas a ribosomas son generadas mediante la traducción *in vitro* de mRNAs (marrón) que carecen de codón de parada. El número mínimo de residuos requerido para abarcar la distancia entre el sitio P del ribosoma y el centro catalítico de la OST ( $d$ ) dependerá del grado de compactación de la región de la cadena nascente situada en el interior del túnel del ribosoma.

*Determinación experimental del grado de interfacialidad de una secuencia: Ensayo 'Lep3G'*

El ensayo Lep3G se ha diseñado también a partir del ensayo 'Lep' (Figura 3), que se ha modificado desplazando el sitio aceptor de glicosilación G2, acercándolo 5 residuos tras el final de la secuencia de estudio. El motivo del acercamiento de la diana G2 es la detección de secuencias dispuestas en la interfase de la membrana: si la secuencia se dispone interfacialmente, un aceptor suficientemente próximo no podrá alcanzar el centro catalítico de la OST y por tanto no será glicosilado (Figura 5, centro). Adicionalmente, se ha añadido una tercera diana (G3) más alejada de la secuencia de estudio. El motivo de esta diana G3 es poder discernir entre una disposición insertada y una interfacial, ya que si la secuencia está dispuesta interfacialmente el largo dominio P2 se encontrará en el lado luminal del microsoma con la oportunidad de glicosilarse. Por el contrario, si la secuencia estudiada se encuentra insertada, únicamente la diana G1 se verá modificada, ya que el dominio P2 estará situado en el lado citosólico de la membrana. Las tres dianas (G1, G2 y G3) únicamente se glicosilarán simultáneamente en el caso que la secuencia de estudio se transloque completamente (Figura 5, derecha).



**Figura 5 | Representación del ensayo Lep3G utilizado para determinar el grado de interfacialidad de una secuencia.**

Esquema del ensayo Lep3G donde tres sitios aceptores de glicosilación (G1, G2 y G3) están estratégicamente distribuidos a lo largo de la construcción. Los rectángulos negros corresponden a las hélices H1 y H2 de Lep, mientras que los rectángulos coloreados corresponden a una disposición final de la secuencia insertada (verde), interfacial (amarillo) o translocada (rojo). Cuando la secuencia de estudio se inserta únicamente se glicosila el sitio aceptor G1 (izquierda). Si por el contrario la secuencia se transloca, los 3 sitios aceptores estarán disponibles para ser glicosilados (derecha). En el caso de que se disponga interfacialmente se modificarán las dianas G1 y G3 (centro).

## Resumen

### *Transcripción y traducción in vitro*

Las construcciones realizadas para los ensayos de ‘Lep’, la determinación del plegamiento en el túnel del ribosoma y los ensayos ‘Lep3G’ fueron transcritas y traducidas *in vitro* utilizando el sistema TNT T7 Quick coupled System (Promega). Las reacciones contenían 5  $\mu\text{L}$  de TNT T7, 1  $\mu\text{L}$  de producto de PCR directamente del mix de amplificación, 0.5  $\mu\text{L}$  de aminoácidos marcados con  $^{35}\text{S}$  (EasyTag EXPRESS  $^{35}\text{S}$  Protein Labeling Mix (5.5  $\mu\text{Ci}$ , Perkin Elmer)) y 0.3  $\mu\text{L}$  de microsomas (tRNA Probes). Las traducciones fueron incubadas a 30 °C durante 30 o 90 minutos, dependiendo de la ausencia o presencia de codón de stop, respectivamente. Las reacciones se detenían añadiendo 50  $\mu\text{L}$  de 1x tampón de carga de proteínas. Las muestras fueron analizadas por electroforesis en gel de poliacrilamida en presencia de SDS (dodecil sulfato sódico) y visualizadas mediante autorradiografía empleando un Fujifilm Image Analyzer modelo FLA-3000. Las bandas de proteínas fueron cuantificadas mediante el software ImageJ (NIH).

### *Tratamientos con Endoglicosidasa H y Proteinasa K*

Para llevar a cabo el tratamiento con *endoglicosidasa H*, una enzima capaz de liberar los *N*-oligosacáridos de una glicoproteína, se toman 18  $\mu\text{L}$  de la reacción de traducción detenida, se mezclan con 2  $\mu\text{L}$  de NEB GlycoBuffer 3 y 2 U de Endoglycosidase H enzyme (New England Biolabs) y se incuba durante 30 min a 37 °C. Tras la incubación, los productos resultantes son analizados mediante electroforesis en gel de poliacrilamida en presencia de SDS y visualizadas mediante autorradiografía.

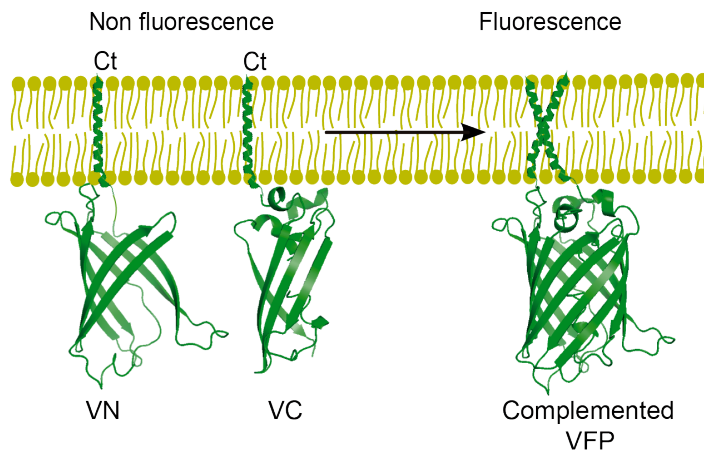
Para llevar a cabo el ensayo de protección de *proteínasa K*, una serín proteasa de amplio espectro que degrada las regiones extramembrana de las proteínas, 5  $\mu\text{L}$  de la mezcla de traducción (sin presencia de tampón de carga de proteínas, para evitar la disrupción de los microsomas por acción de los detergentes) fueron suplementados con 1  $\mu\text{L}$  de proteinasa K (2 mg/mL, Merck) e incubados 20 min en hielo. La reacción de digestión fue detenida añadiendo 120  $\mu\text{L}$  de una solución de 200  $\mu\text{M}$  de fluoruro de fenilmetilsulfonilo (PMSF, Merck) en tampón PBS. El paso siguiente fue recoger las membranas por ultracentrifugación (100,000 x g, 4°C, 15 min) sobre un colchón de sacarosa 0.5 M. Los sobrenadantes fueron cuidadosamente eliminados, resuspendiéndose los microsomas sedimentados en 40  $\mu\text{L}$  de tampón de carga de proteínas. Las muestras fueron analizadas por electroforesis en gel de poliacrilamida en presencia de SDS y visualizadas mediante autorradiografía.



*Ensayo de complementación fluorescente bimolecular: BiFC*

Uno de los ensayos más ampliamente utilizados para el estudio de interacciones entre proteínas es la fluorescencia por complementación bimolecular, también conocida como BiFC. El ensayo se basa en las propiedades de la proteína fluorescente Venus (VFP), la cual se ha partido en dos mitades no fluorescentes (extremo N-terminal (VN), residuos 1-173 y extremo C-terminal (VC), residuos 155-238) que no muestran afinidad intrínseca la una por la otra. Estas dos mitades únicamente son capaces de complementar y recuperar la fluorescencia nativa de la proteína completa cuando son fusionadas a proteínas que interactúan entre sí. De esta forma las dos mitades de la VFP quedan próximas entre sí, permitiendo el plegamiento de la proteína y recuperando su fluorescencia.

Para llevar a cabo el ensayo, ambos plásmidos que codifican para VN y VC se fusionan a la proteína de interés, en nuestro caso TMDs, y se expresan simultáneamente en células eucariotas en cultivo. Las células HEK-293T son transfectadas con 250 ng de cada plásmido VN y VC utilizando PEI como agente de transfección. Adicionalmente se transfecta con 25 ng de un plásmido que codifica para la proteína luciferasa. La luciferasa se emplea meramente como control de transfección y viabilidad de las células. Tras 48 horas de expresión, las células son medidas utilizando un lector multiplaca Victor X4 (Perkin Elmer).



**Figura 6 | Representación esquemática del ensayo BiFC.**

La oligomerización mediada por los segmentos transmembrana (hélices verdes embebidas en la membrana) lleva a la complementación de las dos mitades (extremo N-terminal (VN) y C-terminal (VC)) no fluorescentes de la proteína fluorescente Venus (VFP). Una vez complementadas las dos mitades la fluorescencia queda reconstituida.

## Resumen

### *Dicroísmo circular*

Los péptidos por caracterizar estructuralmente fueron inicialmente resuspendidos en 1 mM de tampón fosfato pH 7.5. Para los ensayos de dicroísmo circular se emplearon mezclas de péptido:lípido (POPC) en una concentración molar de 1:200, dejándose incubar al menos 45 min a pH 7.5. Tras la incubación el pH fue modificado a pH 4 o pH 7 en función del ensayo. La concentración de final de sales en la muestra fue de 5 mM, mientras que la de péptido de 5  $\mu$ M. Las muestras fueron incubadas otros 45 minutos adicionales con los nuevos tampones. Los ensayos de dicroísmo fueron llevados a cabo en un espectropolarímetro Jasco-J815 a 25 °C en una cubeta de 2 mm de grosor. Los datos fueron recogidos cada 1 nm a una velocidad de 50 nm/min en el rango de 260 a 195 nm, con un ancho de banda de 2 nm y un total de 20 muestreos acumulados. La señal de fondo causada por los lípidos fue eliminada de los datos finales. Los valores finales fueron normalizados a miligrados de elipticidad utilizando la siguiente ecuación:

$$\Theta \text{ (deg} \cdot \text{cm}^2 \cdot \text{dmol}^{-1} \cdot 10^3) = \frac{\text{elipticidad (mdeg)} \cdot 10^6}{\text{tamaño cubeta (mm)} \cdot [\text{péptido}] (\mu\text{M}) \cdot n}$$
 donde n es el número de aminoácidos en el péptido menos 1 (número de enlaces peptídicos).

### *Ensayo de liberación de contenidos*

Los lípidos POPC en polvo se rehidrataron con una solución de 20 mM de Sulforrodamina B (S1307, Thermo) y se generaron las vesículas unilamelares uniformes utilizando un extruder con filtros de 200 nm de diámetro de poro. Las vesículas con Sulforrodamina B encapsulada se purificaron mediante columnas de desalado Sephadex G-25 (GE Healthcare) y se eluyeron de tampón fosfato 1 mM (pH 7.5).

Las relaciones molares de péptido:lípido se modificaron manteniendo constante la concentración de vesículas con Sulforrodamina B encapsulada (90  $\mu$ M) y variando la concentración de péptido utilizado de 3 a 500 nM en un volumen final de reacción de 200  $\mu$ L. En ensayo fue llevado a cabo utilizando un lector de fluorescencia BiotekCytation 5 Imaging Reader (Biotek), midiendo el incremento (485/ 590 nm excitación/emisión) producido por la liberación de contenidos de las vesículas. El porcentaje de liberación de contenidos fue calculado utilizando la siguiente ecuación:

$$\% \text{ liberación} = \frac{\text{muestra-control}}{\text{triton-control}} \times 100$$
 donde Triton X-100 fue utilizado como control de liberación total de contenidos, y la muestra sin péptido ni Triton X-100 como muestra sin liberación de contenidos.

## Conclusiones

1. La hidrofobicidad, helicidad y longitud de los segmentos helicoidales son los principales determinantes para la adquisición de la estructura en  $\alpha$ -hélice en el túnel del ribosoma.
2. Se ha desarrollado el ensayo Lep3G que posibilita el estudio de regiones con una disposición interfacial en la membrana de una forma rápida, sistemática y cuantitativa. Este ensayo nos ha permitido a calcular una escala biológica de la interfacialidad de cada uno de los 20 aminoácidos naturales.
3. Las membranas biológicas albergan interacciones homo- y hetero-oligoméricas entre TMDs muy diversos en tamaño, enfatizando la adaptabilidad entre lípidos y péptidos para minimizar el desajuste hidrofóbico.
4. La inserción y topología final de proteínas de membrana no viene sólo marcada por la hidrofobicidad de sus segmentos transmembrana sino también por las interacciones entre ellos y las regiones que los conectan.



## **REFERENCES**



- Abell, B.M., High, S. & Moloney, M.M., 2002. Membrane protein topology of oleosin is constrained by its long hydrophobic domain. *Journal of Biological Chemistry*, 277(10), pp.8602–8610.
- Allen, K.N. et al., 2019. Monotopic Membrane Proteins Join the Fold. *Trends in biochemical sciences*, 44(1), pp.7–20.
- Almén, M.S. et al., 2009. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC biology*, 7, p.50.
- Anbazzhagan, V. & Schneider, D., 2010. The membrane environment modulates self-association of the human GpA TM domain-Implications for membrane protein folding and transmembrane signaling. *Biochimica et biophysica acta*, 1798(10), pp.1899–1907.
- Andersen, O.S. & Koeppe, R.E., 2007. Bilayer thickness and membrane protein function: An energetic perspective. *Annual review of biophysics and biomolecular structure*, 36, pp.107–130.
- Andreev, O.A. et al., 2007. Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo. *Proceedings of the National Academy of Sciences*, 104(19), pp.7893–7898.
- Andreev, O.A. et al., 2010. pH (low) insertion peptide (pHLIP) inserts across a lipid bilayer as a helix and exits by a different path. *Proceedings of the National Academy of Sciences of the United States of America*, 107(9), pp.4081–4086.
- Andreev, O.A., Engelman, D.M. & Reshetnyak, Y.K., 2009. Targeting acidic diseased tissue: New technology based on use of the pH (Low) Insertion Peptide (pHLIP). *Chimica oggi*, 27(2), pp.34–37.
- Andreu-Fernández, V. et al., 2017. Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. In *Proceedings of the National Academy of Sciences of the United States of America*. pp. 310–315.
- Andreu-Fernández, V. et al., 2014. Peptides Derived from the Transmembrane Domain of Bcl-2 Proteins as Potential Mitochondrial Priming Tools. *ACS Chemical Biology*, 9(8), pp.1799–1811.
- Andreu-Fernández, V. et al., 2016. The C-terminal Domains of Apoptotic BH3-only Proteins Mediate Their Insertion into Distinct Biological Membranes. *Journal of Biological Chemistry*, 291(48), pp.25207–25216.
- Baeza-Delgado, C. et al., 2016. Biological insertion of computationally designed short transmembrane segments. *Scientific reports*, 6(1), pp.23397–9.
- Baeza-Delgado, C., Marti-Renom, M.A. & Mingarro, I., 2013. Structure-based statistical analysis of transmembrane helices helices. *European biophysics journal : EBJ*, 42(2-3), pp.199–207.
- Bai, L. et al., 2018. The atomic structure of a eukaryotic oligosaccharyltransferase complex. *Nature Publishing Group*, 555(7696), pp.328–333.
- Balali-Mood, K., Bond, P.J. & Sansom, M.S.P., 2009. Interaction of monotopic membrane enzymes with a lipid bilayer: A coarse-grained MD simulation study. *Biochemistry*, 48(10), pp.2135–2145.

## References

- Ban, N. et al., 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, 289(5481), pp.905–920.
- Bañó-Polo, M. et al., 2020. Insertion of Bacteriorhodopsin Helix C Variants into Biological Membranes. *ACS Omega*, 5(1), pp.556–560.
- Bañó-Polo, M. et al., 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 : a publication of the Protein Society*, 20(1), pp.179–186.
- Bañó-Polo, M. et al., 2012. Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing. *PloS one*, 7(9), p.e44263.
- Basañez, G., Soane, L. & Hardwick, J.M., 2012. A new view of the lethal apoptotic pore. *PLoS biology*, 10(9), p.e1001399.
- Becker, T. et al., 2009. Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science*, 326(5958), pp.1369–1373.
- Beckmann, R. et al., 1997. Alignment of conduits for the nascent polypeptide chain in the ribosome- Sec61 complex. *Science*, 278(5346), pp.2123–2126.
- Beckmann, R. et al., 2001. Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell*, 107(3), pp.361–372.
- Behrmann, E. et al., 2015. Structural snapshots of actively translating human ribosomes. *Cell*, 161(4), pp.845–857.
- Berger, B.W. et al., 2010. Consensus motif for integrin transmembrane helix association. *Proceedings of the National Academy of Sciences of the United States of America*, 107(2), pp.703–708.
- Bhushan, S. et al., 2010. alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nature structural & molecular biology*, 17(3), pp.313–317.
- Bleicken, S. et al., 2014. Structural Model of Active Bax at the Membrane. *Molecular Cell*, 56(4), pp.496–505.
- Bloom, M., Evans, E. & Mouritsen, O.G., 1991. Physical properties of the fluid lipid-bilayer component of cell membranes: A perspective. *Quarterly Reviews of Biophysics*, 24(3), pp.293–397.
- Bond, P.J. & Sansom, M.S.P., 2006. Insertion and assembly of membrane proteins via simulation. *Journal of the American Chemical Society*, 128(8), pp.2697–2704.
- Booth, P.J., 2012. A successful change of circumstance: a transition state for membrane protein folding. *Current opinion in structural biology*, 22(4), pp.469–475.
- Booth, P.J. et al., 1995. Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nature structural & molecular biology*, 2(2), pp.139–143.
- Borgese, N. & Fasana, E., 2011. Targeting pathways of C-tail-anchored proteins. *Biochimica et biophysica acta*, 1808(3), pp.937–946.
- Braunger, K. et al., 2018. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science*, 360(6385), pp.215–219.



- Cannon, K.S. et al., 2005. Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *The Journal of cell biology*, 169(2), pp.219–225.
- Cassaignau, A.M.E., Cabrita, L.D. & Christodoulou, J., 2020. How Does the Ribosome Fold the Proteome? *Annual review of biochemistry*, 89(1), pp.389–415.
- Chamberlain, L.H., Burgoyne, R.D. & Gould, G.W., 2001. SNARE proteins are highly enriched in lipid rafts in PC12 cells: Implications for the spatial control of exocytosis. *Proceedings of the National Academy of Sciences of the United States of America*, 98(10), pp.5619–5624.
- Chao, D.T. & Korsmeyer, S.J., 1998. BCL-2 family: Regulators of cell death, Chitwood, P.J. & Hegde, R.S., 2020. An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature*, 584(7822), pp.630–634.
- Cornell, R.B. & Taneva, S.G., 2006. Amphipathic helices as mediators of the membrane interaction of amphitropic proteins, and as modulators of bilayer physical properties. *Current protein & peptide science*, 7(6), pp.539–552.
- Cosson, P., Perrin, J. & Bonifacino, J.S., 2013. Anchors aweigh: Protein localization and transport mediated by transmembrane domains. *Trends in Cell Biology*, 23(10), pp.511–517.
- Cross, B.C.S. & High, S., 2009. Dissecting the physiological role of selective transmembrane-segment retention at the ER translocon. *Journal of Cell Science*, 122(11), pp.1768–1777.
- Cross, B.C.S. et al., 2009. Delivering proteins for export from the cytosol. *Nature Reviews Molecular Cell Biology*, 10(4), pp.255–264.
- Cymer, F., Heijne, von, G. & White, S.H., 2015. Mechanisms of integral membrane protein insertion and folding. *Journal of molecular biology*, 427(5), pp.999–1022.
- Dale, H., Angevine, C.M. & Krebs, M.P., 2000. Ordered membrane insertion of an archaeal opsin in vivo. *Proceedings of the National Academy of Sciences*, 97(14), pp.7847–7852.
- Dawaliby, R. et al., 2016. Phosphatidylethanolamine Is a Key Regulator of Membrane Fluidity in Eukaryotic Cells. *Journal of Biological Chemistry*, 291(7), pp.3658–3667.
- de Jesus, A.J. & Allen, T.W., 2013. The determinants of hydrophobic mismatch response for transmembrane helices. *Biochimica et biophysica acta*, 1828(2), pp.851–863.
- De Siervo, A.J., 1969. Alterations in the phospholipid composition of *Escherichia coli* B during growth at different temperatures. *Journal of bacteriology*, 100(3), pp.1342–1349.
- DeGrado, W.F., Gratkowski, H. & Lear, J.D., 2003. How do helix-helix interactions help determine the folds of membrane proteins? Perspectives from the study of homo-oligomeric helical bundles. *Protein science : a publication of the Protein Society*, 12(4), pp.647–665.

## References

- Denzer, A.J., Nabholz, C.E. & Spiess, M., 1995. Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO Journal*, 14(24), pp.6311–6317.
- Destainville, N., Schmidt, T.H. & Lang, T., 2016. Where Biology Meets Physics—A Converging View on Membrane Microdomain Dynamics. In *Current Topics in Membranes*. Current Topics in Membranes. Elsevier, pp. 27–65.
- Devaraneni, P.K. et al., 2011. Stepwise insertion and inversion of a type II signal anchor sequence in the ribosome-Sec61 translocon complex. *Cell*, 146(1), pp.134–147.
- Disalvo, E.A. et al., 2008. Structural and functional properties of hydration and confined water in membrane interfaces. *Biochimica et biophysica acta*, 1778(12), pp.2655–2670.
- Doñate-Macian, P. et al., 2015. Molecular and topological membrane folding determinants of transient receptor potential vanilloid 2 channel. *Biochemical and Biophysical Research Communications*, 462(3), pp.221–226.
- Dowhan, W., Vitrac, H. & Bogdanov, M., 2019. Lipid-Assisted Membrane Protein Folding and Topogenesis. *Protein Journal*, 38(3), pp.274–288.
- Dupuy, A.D. & Engelman, D.M., 2008. Protein area occupancy at the center of the red blood cell membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8), pp.2848–2852.
- Egea, P.F. & Stroud, R.M., 2010. Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proceedings of the National Academy of Sciences of the United States of America*, 107(40), pp.17182–17187.
- Eisenberg, D., Weiss, R.M. & Terwilliger, T.C., 1982. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, 299(5881), pp.371–374.
- Engelman, D.M., 2005. Membranes are more mosaic than fluid. *Nature*, 438(7068), pp.578–580.
- Engelman, D.M. et al., 2003. Membrane protein folding: beyond the two stage model. *FEBS Letters*, 555(1), pp.122–125.
- Epand, R.F. et al., 2002. Direct evidence for membrane pore formation by the apoptotic protein Bax. *Biochemical and Biophysical Research Communications*, 298(5), pp.744–749.
- Estrella, V. et al., 2013. Acidity generated by the tumor microenvironment drives local invasion. *Cancer Research*, 73(5), pp.1524–1535.
- Faham, S. et al., 2004. Side-chain contributions to membrane protein structure and stability. *Journal of molecular biology*, 335(1), pp.297–305.
- Fariás-Rico, J.A. et al., 2017. Mutational analysis of protein folding inside the ribosome exit tunnel. *FEBS Letters*, 591(1), pp.155–163.
- Fjell, C.D., Hancock, R.E.W. & Cherkasov, A., 2007. AMPper: a database and an automated discovery tool for antimicrobial peptides. *Bioinformatics*, 23(9), pp.1148–1155.

- Flinner, N., Mirus, O. & Schleiff, E., 2014. The influence of fatty acids on the GpA dimer interface by coarse-grained molecular dynamics simulation. *International journal of molecular sciences*, 15(8), pp.14247–14268.
- Fons, R.D., Bogert, B.A. & Hegde, R.S., 2003. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *Journal of Cell Biology*, 160(4), pp.529–539.
- Frank, J. et al., 1995. A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature*, 376(6539), pp.441–444.
- Garcia-Saez, A.J. et al., 2006. Peptides corresponding to helices 5 and 6 of Bax can independently form large lipid pores. *The FEBS journal*, 273(5), pp.971–981.
- Garrett, R.A. et al., 2000. *The Ribosome*, Washington: American Society for Microbiology.
- Gautier, R. et al., 2008. HELIQUEST: a web server to screen sequences with specific alpha-helical properties. *Bioinformatics*, 24(18), pp.2101–2102.
- Gavel, Y. & Heijne, G.V., 1990. Sequence differences between glycosylated and non-glycosylated asn-x-thr/ser acceptor sites: Implications for protein engineering. *Protein Engineering, Design and Selection*, 3(5), pp.433–442.
- Giménez-Andrés, M., Čopič, A. & Antonny, B., 2018. The many faces of amphipathic helices. *Biomolecules*, 8(3).
- Goder, V. & Spiess, M., 2003. Molecular mechanism of signal sequence orientation in the endoplasmic reticulum. *EMBO Journal*, 22(14), pp.3645–3653.
- Goder, V. & Spiess, M., 2001. Topogenesis of membrane proteins: Determinants and dynamics. In *FEBS Letters*. pp. 87–93.
- Goder, V., Junne, T. & Spiess, M., 2004. Sec61p Contributes to Signal Sequence Orientation According to the Positive-Inside Rule. *Molecular Biology of the Cell*, 15(3), pp.1470–1478.
- Gogala, M. et al., 2014. Structures of the Sec61 complex engaged in nascent peptide translocation or membrane insertion. *Nature*, 506(7486), pp.107–110.
- Görllich, D. & Rapoport, T.A., 1993. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell*, 75(4), pp.615–630.
- Grau, B. et al., 2017. The role of hydrophobic matching on transmembrane helix packing in cells. *Cell Stress*, 1(2), pp.90–106.
- Grau-Campistany, A. et al., 2016. Extending the Hydrophobic Mismatch Concept to Amphiphilic Membranolytic Peptides. *Journal of Physical Chemistry Letters*, 7(7), pp.1116–1120.
- Gumbart, J. et al., 2012. Mechanisms of SecM-mediated stalling in the ribosome. *Biophysical journal*, 103(2), pp.331–341.
- Gumbart, J., Chipot, C. & Schulten, K., 2011. Free energy of nascent-chain folding in the translocon. *Journal of the American Chemical Society*, 133(19), pp.7602–7607.
- Hamman, B.D. et al., 1997. The aqueous pore through the translocon has a diameter of 40–60 Å during cotranslational protein translocation at the ER membrane. *Cell*, 89(4), pp.535–544.

## References

- Harayama, T. & Riezman, H., 2018. Understanding the diversity of membrane lipid composition. *Nature Reviews Molecular Cell Biology*, 19(5), pp.281–296.
- Hartmann, E. et al., 1993. A tetrameric complex of membrane proteins in the endoplasmic reticulum. *European journal of biochemistry*, 214(2), pp.375–381.
- Hartmann, E., Rapoport, T.A. & Lodish, H.F., 1989. Predicting the orientation of eukaryotic membrane-spanning proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 86(15), pp.5786–5790.
- Hedman, R. et al., 2015. Cotranslational Protein Folding inside the Ribosome Exit Tunnel. *Cell Reports*, 12(10), pp.1533–1540.
- Hegde, R.S., Voigt, S. & Lingappa, V.R., 1998. Regulation of protein topology by trans-acting factors at the endoplasmic reticulum. *Molecular Cell*, 2(1), pp.85–91.
- Heijne, von, G., 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. *EMBO Journal*, 3(10), pp.2315–2318.
- Heijne, von, G., 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*, 341(6241), pp.456–458.
- Heijne, von, G., 2006. Membrane-protein topology. *Nature Publishing Group*, 7(12), pp.909–918.
- Heijne, von, G., 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO Journal*, 5(11), pp.3021–3027.
- Heinrich, S.U. et al., 2000. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell*, 102(2), pp.233–244.
- Henry, K.E. et al., 2020. Demarcation of Sepsis-Induced Peripheral and Central Acidosis with pH-Low Insertion Cyclic (pHLIC) Peptide. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, p.jnumed.119.233072.
- Hessa, T. et al., 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450(7172), pp.1026–1030.
- Hessa, T. et al., 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*, 433(7024), pp.377–381.
- Hénin, J., Pohorille, A. & Chipot, C., 2005. Insights into the recognition and association of transmembrane  $\alpha$ -helices. The free energy of  $\alpha$ -helix dimerization in glycophorin A. *Journal of the American Chemical Society*, 127(23), pp.8478–8484.
- Hilger, D., Masureel, M. & Kobilka, B.K., 2018. Structure and dynamics of GPCR signaling complexes. *Nature structural & molecular biology*, 25(1), pp.4–12.
- Hong, H., 2014. Toward understanding driving forces in membrane protein folding. *Archives of Biochemistry and Biophysics*, 564, pp.297–313.

- Hristova, K., Dempsey, C.E. & White, S.H., 2001. Structure, location, and lipid perturbations of melittin at the membrane interface. *Biophysical journal*, 80(2), pp.801–811.
- Hunt, J.F. et al., 1997. A biophysical study of integral membrane protein folding. *Biochemistry*, 36(49), pp.15156–15176.
- Janosi, L., Prakash, A. & Doxastakis, M., 2010. Lipid-modulated sequence-specific association of glycophorin a in membranes. *Biophysical journal*, 99(1), pp.284–292.
- Janovjak, H. et al., 2003. Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO Journal*, 22(19), pp.5220–5229.
- Jaud, S. et al., 2009. Insertion of short transmembrane helices by the Sec61 translocon. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), pp.11588–11593.
- Javadpour, M.M. et al., 1999. Helix packing in polytopic membrane proteins: Role of glycine in transmembrane helix association. *Biophysical journal*, 77(3), pp.1609–1618.
- Jayasinghe, S., Hristova, K. & White, S.H., 2001. Energetics, stability, and prediction of transmembrane helices. *Journal of molecular biology*, 312(5), pp.927–934.
- Jensen, M. & Mouritsen, O.G., 2004. Lipids do influence protein function - The hydrophobic matching hypothesis revisited. *Biochimica et biophysica acta*, 1666(1-2), pp.205–226.
- Jiang, Y. et al., 2008. An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *Journal of Cell Biology*, 180(6), pp.1149–1161.
- Joh, N.H. et al., 2008. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature Publishing Group*, 453(7199), pp.1266–1270.
- Johansson, A.C.Y. & Lindahl, E., 2009. Titratable amino acid solvation in lipid membranes as a function of protonation state. *Journal of Physical Chemistry B*, 113(1), pp.245–253.
- Johnson, A.E. & van Waes, M.A., 1999. The translocon: a dynamic gateway at the ER membrane. *Annual Review of Cell and Developmental Biology*, 15(1), pp.799–842.
- Junne, T. et al., 2007. Mutations in the Sec61p channel affecting signal sequence recognition and membrane protein topology. *Journal of Biological Chemistry*, 282(45), pp.33201–33209.
- Junne, T., Kocik, L. & Spiess, M., 2010. The hydrophobic core of the Sec61 translocon defines the hydrophobicity threshold for membrane integration. R. Gilmore, ed. *Molecular Biology of the Cell*, 21(10), pp.1662–1670.
- Kahn, T.W. & Engelman, D.M., 1992. Bacteriorhodopsin can be refolded from two independently stable transmembrane helices and the complementary five-helix fragment. *Biochemistry*, 31(26), pp.6144–6151.
- Kalies, K.U., Rapoport, T.A. & Hartmann, E., 1998. The  $\beta$  subunit of the Sec61 complex facilitates cotranslational protein transport and interacts with the

## References

- signal peptidase during translocation. *Journal of Cell Biology*, 141(4), pp.887–894.
- Kandasamy, S.K. & Larson, R.G., 2006. Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch. *Biophysical journal*, 90(7), pp.2326–2343.
- Kelly, S.M. & Price, N.C., 2000. The use of circular dichroism in the investigation of protein structure and function. *Current protein & peptide science*, 1(4), pp.349–384.
- Kelly, S.M., Jess, T.J. & Price, N.C., 2005. How to study proteins by circular dichroism. *Biochimica et biophysica acta*, 1751(2), pp.119–139.
- Kida, Y., Mihara, K. & Sakaguchi, M., 2005. Translocation of a long amino-terminal domain through ER membrane by following signal-anchor sequence. *EMBO Journal*, 24(18), pp.3202–3213.
- Killian, J.A., 1998. Hydrophobic mismatch between proteins and lipids in membranes. *Biochimica et biophysica acta*, 1376(3), pp.401–416.
- Kim, S. et al., 2005. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proceedings of the National Academy of Sciences*, 102(40), pp.14278–14283.
- Kim, T. & Im, W., 2010. Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. *Biophysical journal*, 99(1), pp.175–183.
- Kokubo, H. & Okamoto, Y., 2009. Analysis of helix-helix interactions of bacteriorhodopsin by replica-exchange simulations. *Biophysical journal*, 96(3), pp.765–776.
- Kosolapov, A. & Deutsch, C., 2009. Tertiary interactions within the ribosomal exit tunnel. *Nature structural & molecular biology*, 16(4), pp.405–411.
- Krishnamani, V. & Lanyi, J.K., 2012. Molecular dynamics simulation of the unfolding of individual bacteriorhodopsin helices in sodium dodecyl sulfate micelles. *Biochemistry*, 51(6), pp.1061–1069.
- Kuhlman, B. et al., 1997. An exceptionally stable helix from the ribosomal protein L9: implications for protein folding and stability. *Journal of molecular biology*, 270(5), pp.640–647.
- Kuwana, T. et al., 2002. Bid, Bax, and Lipids Cooperate to Form Supramolecular Openings in the Outer Mitochondrial Membrane. *Cell*, 111(3), pp.331–342.
- Kuznetsov, A.S., Volynsky, P.E. & Efremov, R.G., 2015. Role of the Lipid Environment in the Dimerization of Transmembrane Domains of Glycophorin A. *Acta naturae*, 7(4), pp.122–127.
- Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), pp.860–921.
- Lang, T., 2007. SNARE proteins and 'membrane rafts'. *The Journal of physiology*, 585(Pt 3), pp.693–698.
- Langosch, D. et al., 1996. Dimerisation of the glycophorin A transmembrane segment in membranes probed with the ToxR transcription activator. *Journal of molecular biology*, 263(4), pp.525–530.

- Lee, A.G., 2005. How lipids and proteins interact in a membrane: A molecular approach. *Molecular BioSystems*, 1(3), pp.203–212.
- Lemmon, M.A. et al., 1992. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *Journal of Biological Chemistry*, 267(11), pp.7683–7689.
- Li, R. et al., 2004. Dimerization of the transmembrane domain of Integrin alphaIIb subunit in cell membranes. *Journal of Biological Chemistry*, 279(25), pp.26666–26673.
- Liao, S. et al., 1997. Both lumenal and cytosolic gating of the aqueous ER translocon pore are regulated from inside the ribosome during membrane protein integration. *Cell*, 90(1), pp.31–41.
- Liljas, A. et al., 2009. *Textbook of Structural Biology*, World Scientific Publishing Company.
- Lin, P.-J. et al., 2011. Polytopic membrane protein folding at L17 in the ribosome tunnel initiates cyclical changes at the translocon. *Journal of Cell Biology*, 195(1), pp.55–70.
- Lomize, A.L. et al., 2007. The role of hydrophobic interactions in positioning of peripheral proteins in membranes. *BMC Structural Biology*, 7, p.44.
- Lomize, M.A. et al., 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic acids research*, 40(Database issue), pp.D370–6.
- Longo, P.A. et al., 2013. Transient Mammalian Cell Transfection with Polyethylenimine (PEI). In *Laboratory Methods in Enzymology: DNA. Methods in Enzymology*. Elsevier, pp. 227–240.
- Lu, J. & Deutsch, C., 2005a. Folding zones inside the ribosomal exit tunnel. *Nature structural & molecular biology*, 12(12), pp.1123–1129.
- Lu, J. & Deutsch, C., 2005b. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry*, 44(23), pp.8230–8243.
- Luecke, H. et al., 1999. Structure of bacteriorhodopsin at 1.55 Å resolution. *Journal of molecular biology*, 291(4), pp.899–911.
- Lundin, C. et al., 2008. Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N(in)-C(out) and N(out)-C(in) transmembrane helices. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), pp.15702–15707.
- Lycklama a Nijeholt, J.A., Wu, Z.C. & Driessen, A.J.M., 2011. Conformational dynamics of the plug domain of the SecYEG protein-conducting channel. *Journal of Biological Chemistry*, 286(51), pp.43881–43890.
- MacCallum, J.L. & Tieleman, D.P., 2011. Hydrophobicity scales: a thermodynamic looking glass into lipid-protein interactions. *Trends in biochemical sciences*, 36(12), pp.653–662.
- MacKenzie, K.R., 2006. Folding and stability of  $\alpha$ -helical integral membrane proteins. *Chemical Reviews*, 106(5), pp.1931–1977.
- MacKenzie, K.R., Prestegard, J.H. & Engelman, D.M., 1997. A transmembrane helix dimer: structure and implications. *Science*, 276(5309), pp.131–133.

## References

- Malkin, L.I. & Rich, A., 1967. Partial resistance of nascent polypeptide chains to proteolytic digestion due to ribosomal shielding. *Journal of molecular biology*, 26(2), pp.329–346.
- Marti, T., 1998. Refolding of bacteriorhodopsin from expressed polypeptide fragments. *Journal of Biological Chemistry*, 273(15), pp.9312–9322.
- Martínez-Gil, L. & Mingarro, I., 2015. Viroporins, examples of the two-stage membrane protein folding model. *Viruses*, 7(7), pp.3462–3482.
- Martínez-Gil, L. et al., 2011. Membrane protein integration into the endoplasmic reticulum. *FEBS Journal*, 278(20), pp.3846–3858.
- Martínez-Gil, L., Johnson, A.E. & Mingarro, I., 2010. Membrane insertion and biogenesis of the Turnip crinkle virus p9 movement protein. *Journal of virology*, 84(11), pp.5520–5527.
- Mattevi, A., 2006. A monotopic membrane protein goes solo. *Structure (London, England : 1993)*, 14(4), pp.628–629.
- Meacock, S.L. et al., 2002. Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein R. Gilmore, ed. *Molecular Biology of the Cell*, 13(12), pp.4114–4129.
- Ménétrez, J.-F. et al., 2008. Single copies of Sec61 and TRAP associate with a nontranslating mammalian ribosome. *Structure (London, England : 1993)*, 16(7), pp.1126–1137.
- Milovanovic, D. et al., 2015. Hydrophobic mismatch sorts SNARE proteins into distinct membrane domains. *Nature communications*, 6(1), p.720.
- Mingarro, I. et al., 2000. Different conformations of nascent polypeptides during translocation across the ER membrane. *BMC cell biology*, 1(1), pp.3–8.
- Mitra, K. et al., 2004. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proceedings of the National Academy of Sciences*, 101(12), pp.4083–4088.
- Monné, M. et al., 1999. Turns in transmembrane helices: Determination of the minimal length of a “Helical hairpin” and derivation of a fine-grained turn propensity scale. *Journal of molecular biology*, 293(4), pp.807–814.
- Mouritsen, O.G. & Bloom, M., 1984. Mattress model of lipid-protein interactions in membranes. *Biophysical journal*, 46(2), pp.141–153.
- Muhle-Goll, C. et al., 2012. Hydrophobic matching controls the tilt and stability of the dimeric platelet-derived growth factor receptor (PDGFR)  $\alpha$  transmembrane segment. *Journal of Biological Chemistry*, 287(31), pp.26178–26186.
- Musial-Siwek, M. et al., 2010. Tuning the insertion properties of pHLIP. *Biochimica et biophysica acta*, 1798(6), pp.1041–1046.
- Müller, D.J. et al., 2002. Stability of bacteriorhodopsin  $\alpha$ -helices and loops analyzed by single-molecule force spectroscopy. *Biophysical journal*, 83(6), pp.3578–3588.
- Nagasawa, K. et al., 2007. Simultaneous induction of the four subunits of the TRAP complex by ER stress accelerates ER degradation. *EMBO Reports*, 8(5), pp.483–489.



- Nezil, F.A. & Bloom, M., 1992. Combined influence of cholesterol and synthetic amphiphilic peptides upon bilayer thickness in model membranes. *Biophysical journal*, 61(5), pp.1176–1183.
- Ng, B.G. et al., 2015. Expanding the Molecular and Clinical Phenotype of SSR4-CDG. *Human Mutation*, 36(11), pp.1048–1051.
- Nilsson, I., Johnson, A.E. & Heijne, von, G., 2003. How hydrophobic is alanine? *Journal of Biological Chemistry*, 278(32), pp.29389–29393.
- Nilsson, I.M. & Heijne, von, G., 1993. Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *Journal of Biological Chemistry*, 268(8), pp.5798–5801.
- Nissen, P. et al., 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481), pp.920–930.
- Orzáez, M. et al., 2005. Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycoporphin A dimerisation motif. *FEBS Letters*, 579(7), pp.1633–1638.
- Orzáez, M., Pérez-Payá, E. & Mingarro, I., 2000. Influence of the C-terminus of the glycoporphin A transmembrane fragment on the dimerization process. *Protein Science*, 9(6), pp.1246–1253.
- Pang, A.J. & Reithmeier, R.A.F., 2009. Interaction of anion exchanger 1 and glycoporphin A in human erythroleukaemic K562 cells. *Biochemical Journal*, 421(3), pp.345–356.
- Pataki, C.I. et al., 2018. Proteomic analysis of monolayer-integrated proteins on lipid droplets identifies amphipathic interfacial  $\alpha$ -helical membrane anchors. *Proceedings of the National Academy of Sciences of the United States of America*, 115(35), pp.E8172–E8180.
- Peiró, A. et al., 2014. The Tobacco mosaic virus movement protein associates with but does not integrate into biological membranes. *Journal of virology*, 88(5), pp.3016–3026.
- Petrache, H.I. et al., 2000. Modulation of glycoporphin A transmembrane helix interactions by lipid bilayers: Molecular dynamics calculations. *Journal of molecular biology*, 302(3), pp.727–746.
- Pfeffer, S. et al., 2017. Dissecting the molecular organization of the translocon-associated protein complex. *Nature communications*, 8(1), p.14516.
- Pfeffer, S. et al., 2016. Organization of the native ribosome–translocon complex at the mammalian endoplasmic reticulum membrane. *Biochimica et biophysica acta*, 1860(10), pp.2122–2129.
- Pfeffer, S. et al., 2014. Structure of the mammalian oligosaccharyl-transferase complex in the native ER protein translocon. *Nature communications*, 5(1), pp.3072–8.
- Pfeffer, S. et al., 2015. Structure of the native Sec61 protein-conducting channel. *Nature communications*, 6(1), pp.8403–7.
- Pike, L.J., 2006. Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function. In *Journal of lipid research*. pp. 1597–1598.

## References

- Pitonzo, D. et al., 2009. Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum sec61 translocon P. Walter, ed. *Molecular Biology of the Cell*, 20(2), pp.685–698.
- Plath, K. et al., 1998. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell*, 94(6), pp.795–807.
- Plessis, du, D.J.F. et al., 2009. The lateral gate of SecYEG opens during protein translocation. *Journal of Biological Chemistry*, 284(23), pp.15805–15814.
- Pool, M.R., 2009. A trans-membrane segment inside the ribosome exit tunnel triggers RAMP4 recruitment to the Sec61p translocase. *The Journal of cell biology*, 185(5), pp.889–902.
- Pool, M.R., 2005. Signal recognition particles in chloroplasts, bacteria, yeast and mammals (review). *Molecular Membrane Biology*, 22(1-2), pp.3–15.
- Popot, J.-L., Gerchman, S.E. & Engelman, D.M., 1987. Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. *Journal of molecular biology*, 198(4), pp.655–676.
- Popot, J.L. & Engelman, D.M., 2000. Helical membrane protein folding, stability, and evolution. *Annual review of biochemistry*, 69(1), pp.881–922.
- Popot, J.L. & Engelman, D.M., 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, 29(17), pp.4031–4037.
- Quinn, P., Griffiths, G. & Warren, G., 1984. Density of newly synthesized plasma membrane proteins in intracellular membranes II. *Biochemical studies. Journal of Cell Biology*, 98(6), pp.2142–2147.
- Ramadurai, S. et al., 2010. Influence of hydrophobic mismatch and amino acid composition on the lateral diffusion of transmembrane peptides. *Biophysical journal*, 99(5), pp.1447–1454.
- Ramón-Maiques, S. et al., 2006. Structural bases of feed-back control of arginine biosynthesis, revealed by the structures of two hexameric N-acetylglutamate kinases, from *Thermotoga maritima* and *Pseudomonas aeruginosa*. *Journal of molecular biology*, 356(3), pp.695–713.
- Rapoport, T.A., 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, 450(7170), pp.663–669.
- Rathinakumar, R., Walkenhorst, W.F. & Wimley, W.C., 2009. Broad-Spectrum Antimicrobial Peptides by Rational Combinatorial Design and High-Throughput Screening: The Importance of Interfacial Activity. *Journal of the American Chemical Society*, 131(22), pp.7609–7617.
- Ren, J. et al., 1999. Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. *Biochemistry*, 38(18), pp.5905–5912.
- Russ, W.P. & Engelman, D.M., 1999. TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proceedings of the National Academy of Sciences*, 96(3), pp.863–868.
- Sadlish, H. et al., 2005. Sequential triage of transmembrane segments by Sec61 $\alpha$  during biogenesis of a native multispansing membrane protein. *Nature structural & molecular biology*, 12(10), pp.870–878.

- Saurí, A. et al., 2005. Double-spanning plant viral movement protein integration into the endoplasmic reticulum membrane is signal recognition particle-dependent, translocon-mediated, and concerted. *Journal of Biological Chemistry*, 280(27), pp.25907–25912.
- Saurí, A. et al., 2007. Sec61 $\alpha$  and TRAM are Sequentially Adjacent to a Nascent Viral Membrane Protein during its ER Integration. *Journal of molecular biology*, 366(2), pp.366–374.
- Saurí, A. et al., 2009. Viral Membrane Protein Topology Is Dictated by Multiple Determinants in Its Sequence. *Journal of molecular biology*, 387(1), pp.113–128.
- Sääf, A., Wallin, E. & Heijne, von, G., 1998. Stop-transfer function of pseudo-random amino acid segments during translocation across prokaryotic and eukaryotic membranes. *European journal of biochemistry*, 251(3), pp.821–829.
- Scott, H.L., Westerfield, J.M. & Barrera, F.N., 2017. Determination of the Membrane Translocation pK of the pH-Low Insertion Peptide. *Biophysical journal*, 113(4), pp.869–879.
- Sengupta, D. & Marrink, S.J., 2010. Lipid-mediated interactions tune the association of glycophorin A helix and its disruptive mutants in membranes. *Physical Chemistry Chemical Physics*, 12(40), pp.12987–12996.
- Senoune, S.R., Luo, D. & Martínez-Zaguilán, R., 2004. Plasmalemmal vacuolar-type H<sup>+</sup>-ATPase in cancer biology. *Cell biochemistry and biophysics*, 40(2), pp.185–206.
- Seppälä, S. et al., 2010. Control of membrane protein topology by a single C-terminal residue. *Science*, 328(5986), pp.1698–1700.
- Sezgin, E. et al., 2017. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nature Reviews Molecular Cell Biology*, 18(6), pp.361–374.
- Shao, S. & Hegde, R.S., 2011. Membrane protein insertion at the endoplasmic reticulum. *Annual Review of Cell and Developmental Biology*, 27, pp.25–56.
- Sharpe, H.J., Stevens, T.J. & Munro, S., 2010. A Comprehensive Comparison of Transmembrane Domains Reveals Organelle-Specific Properties. *Cell*, 142(1), pp.158–169.
- Simons, K. & Ikonen, E., 1997. Functional rafts in cell membranes. *Nature*, 387(6633), pp.569–572.
- Singer, S.J. & Nicolson, G.L., 1972. The fluid mosaic model of the structure of cell membranes. *Science*, 175(4023), pp.720–731.
- Singh, S. & Mittal, A., 2016. Transmembrane Domain Lengths Serve as Signatures of Organismal Complexity and Viral Transport Mechanisms. *Nature Publishing Group*, 6(1), p.22352.
- Smith, S.O. et al., 2001. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry*, 40(22), pp.6553–6558.
- Sommer, N. et al., 2013. TRAP assists membrane protein topogenesis at the mammalian ER membrane. *Biochimica et biophysica acta*, 1833(12), pp.3104–3111.

## References

- Song, W. et al., 2000. Role of Sec61alpha in the regulated transfer of the ribosome-nascent chain complex from the signal recognition particle to the translocation channel. *Cell*, 100(3), pp.333–343.
- Soubias, O. et al., 2015. Rhodopsin/lipid hydrophobic matching - Rhodopsin oligomerization and function. *Biophysical journal*, 108(5), pp.1125–1132.
- Sparr, E. et al., 2005. Self-association of transmembrane alpha-helices in model membranes: importance of helix orientation and role of hydrophobic mismatch. *Journal of Biological Chemistry*, 280(47), pp.39324–39331.
- Stachura, S. & Kneller, G.R., 2015. Communication: Probing anomalous diffusion in frequency space. *The Journal of chemical physics*, 143(19), p.191103.
- Stefanovic, S. & Hegde, R.S., 2007. Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell*, 128(6), pp.1147–1159.
- Strandberg, E. et al., 2012. Hydrophobic mismatch of mobile transmembrane helices: Merging theory and experiments. *Biochimica et biophysica acta*, 1818(5), pp.1242–1249.
- Su, P.-C. & Berger, B.W., 2013. A novel assay for assessing juxtamembrane and transmembrane domain interactions important for receptor heterodimerization. *Journal of molecular biology*, 425(22), pp.4652–4658.
- Su, T. et al., 2017. The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling. *eLife*.
- Takamori, S. et al., 2006. Molecular Anatomy of a Trafficking Organelle. *Cell*, 127(4), pp.831–846.
- Tamborero, S. et al., 2011. Membrane Insertion and Topology of the Translocating Chain-Associating Membrane Protein (TRAM). *Journal of molecular biology*, 406(4), pp.571–582.
- Tu, L., Khanna, P. & Deutsch, C., 2014. Transmembrane segments form tertiary hairpins in the folding vestibule of the ribosome. *Journal of molecular biology*, 426(1), pp.185–198.
- Tu, L.W. & Deutsch, C., 2010. A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation. *Journal of molecular biology*, 396(5), pp.1346–1360.
- Tu, L.W. & Deutsch, C., 2017. Determinants of Helix Formation for a Kv1.3 Transmembrane Segment inside the Ribosome Exit Tunnel. *Journal of molecular biology*, 429(11), pp.1722–1732.
- Uhlén, M. et al., 2015. Tissue-based map of the human proteome. *Science*, 347(6220), p.1260419.
- Vallone, B. et al., 1998. Free energy of burying hydrophobic residues in the interface between protein subunits. *Proceedings of the National Academy of Sciences*, 95(11), pp.6103–6107.
- Van den Berg, B. et al., 2004. X-ray structure of a protein-conducting channel. *Nature*, 427(6969), pp.36–44.
- van Meer, G., Voelker, D.R. & Feigenson, G.W., 2008. Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2), pp.112–124.

- Venter, J.C. et al., 2001. The sequence of the human genome. *Science*, 291(5507), pp.1304–1351.
- Voorhees, R.M. & Hegde, R.S., 2016. Structure of the Sec61 channel opened by a signal sequence. *Science*, 351(6268), pp.88–89.
- Voorhees, R.M. & Hegde, R.S., 2015. Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. *eLife*.
- Voorhees, R.M. et al., 2014. Structure of the mammalian ribosome-Sec61 complex to 3.4 Å resolution. *Cell*, 157(7), pp.1632–1643.
- Voss, N.R. et al., 2006. The geometry of the ribosomal polypeptide exit tunnel. *Journal of molecular biology*, 360(4), pp.893–906.
- Walters, R.F.S. & DeGrado, W.F., 2006. Helix-packing motifs in membrane proteins. *Proceedings of the National Academy of Sciences*, 103(37), pp.13658–13663.
- Wang, B. et al., 2008. BAP31 Interacts with Sec61 Translocons and Promotes Retrotranslocation of CFTR $\Delta$ F508 via the Derlin-1 Complex. *Cell*, 133(6), pp.1080–1092.
- Wei, P. et al., 2011. The dimerization interface of the glycoprotein Ib $\beta$  transmembrane domain corresponds to polar residues within a leucine zipper motif. *Protein science : a publication of the Protein Society*, 20(11), pp.1814–1823.
- Weiss, T.M. et al., 2003. Hydrophobic mismatch between helices and lipid bilayers. *Biophysical journal*, 84(1), pp.379–385.
- Westphal, D. et al., 2014. Apoptotic pore formation is associated with in-plane insertion of Bak or Bax central helices into the mitochondrial outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 111(39), pp.E4076–85.
- White, S.H. & Heijne, von, G., 2008. How translocons select transmembrane helices. *Annu Rev Biophys*, 37, pp.23–42.
- White, S.H. & Wimley, W.C., 1998. Hydrophobic interactions of peptides with membrane interfaces. *Biochimica et biophysica acta*, 1376(3), pp.339–352.
- White, S.H. & Wimley, W.C., 1999. Membrane protein folding and stability: physical principles. *Annual review of biophysics and biomolecular structure*, 28(1), pp.319–365.
- Whitley, P. & Mingarro, I., 2014. Stitching proteins into membranes, not sew simple. *Biological chemistry*, 395(12), pp.1417–1424.
- Whitley, P., Nilsson, I.M. & Heijne, von, G., 1996. A nascent secretory protein may traverse the ribosome/endoplasmic reticulum translocase complex as an extended chain. *Journal of Biological Chemistry*, 271(11), pp.6241–6244.
- Wiener, M.C. & White, S.H., 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophysical journal*, 61(2), pp.434–447.
- Wijesinghe, D. et al., 2011. Tuning a polar molecule for selective cytoplasmic delivery by a pH (Low) insertion peptide. *Biochemistry*, 50(47), pp.10215–10222.

## References

- Wild, R. et al., 2018. Structure of the yeast oligosaccharyltransferase complex gives insight into eukaryotic N-glycosylation. *Science*, 359(6375), pp.545–550.
- Wilkinson, B.M., Critchley, A.J. & Stirling, C.J., 1996. Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. *Journal of Biological Chemistry*, 271(41), pp.25590–25597.
- Wimley, W.C., 2010. Describing the mechanism of antimicrobial peptide action with the interfacial activity model. *ACS Chemical Biology*, 5(10), pp.905–917.
- Wimley, W.C. & White, S.H., 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature structural biology*, 3(10), pp.842–848.
- Woolhead, C.A., McCormick, P.J. & Johnson, A.E., 2004. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell*, 116(5), pp.725–736.
- Wu, X., Cabanos, C. & Rapoport, T.A., 2019. Structure of the post-translational protein translocation machinery of the ER membrane. *Nature*, 566(7742), pp.136–139.
- Xiao, K. et al., 2016. Alpha 5/6 helix domains together with N-terminus determine the apoptotic potency of the Bcl-2 family proteins. *Apoptosis*, 21(11), pp.1214–1226.
- Yamaguchi, A. et al., 1999. Stress-associated endoplasmic reticulum protein 1 (SERP1)/Ribosome-associated membrane protein 4 (RAMP4) stabilizes membrane proteins during stress and facilitates subsequent glycosylation. *The Journal of cell biology*, 147(6), pp.1195–1204.
- Zhang, G. & Ignatova, Z., 2011. Folding at the birth of the nascent chain: coordinating translation with co-translational folding. *Current opinion in structural biology*, 21(1), pp.25–31.
- Zilly, F.E. et al., 2011. Ca<sup>2+</sup> induces clustering of membrane proteins in the plasma membrane via electrostatic interactions. *EMBO Journal*, 30(7), pp.1209–1220.
- Ziv, G., Haran, G. & Thirumalai, D., 2005. Ribosome exit tunnel can entropically stabilize alpha-helices. *Proceedings of the National Academy of Sciences*, 102(52), pp.18956–18961.

## ANNEX I

**Truncated constructs analyzed in Chapter 1.** Amino acid sequences bridging the distance between the glycosylation acceptor site (N<sup>1</sup>ST) and the end of the translated region for distance 73 (d=73) are shown in single letter code. The last residue for truncates with 67 residues P-NST (d=67) is indicated in each construct with an arrow (↵). Hydrophobic TMD and soluble helical sequences are highlighted in yellow and gray boxes, respectively. Charged residues in these gray boxed regions are shown in red (Asp and Glu) or in blue (Lys and Arg).





**VSV-G construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMITSS IASFFFIIGL IIGLFLVLMH RLSEKKE<sup>67</sup>↓TLG  
 DVT<sup>73</sup>...

**gp41 construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMIKLF IMIVGGLVGL RIVFAVLSVV HMRLSER<sup>67</sup>↓KET  
 LGD<sup>73</sup>...

**NAGK construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMISRD DAAQVAKVLS EALPYIRRFV HMRLSER<sup>67</sup>↓KET  
 LGD<sup>73</sup>...

**rib. L9 construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMIKAL EAQKQKEQRQ AAEEELANAKK HMRLSER<sup>67</sup>↓KET  
 LGD<sup>73</sup>...

**GpA construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMITLI IFGVMAGVIG TILLISYGIK KKKHMRL<sup>67</sup>↓SER  
 KET<sup>73</sup>...

**M13 construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMISYI GYAWAMVVVI VGATIGIKLF KHMRLSE<sup>67</sup>↓RKE  
 TLG<sup>73</sup>...

**p75 construct**

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
 FFEVPMINLI  
 PVYCSILAAV VVGLVAYIAF KRWNHMR<sup>67</sup>↓LSE RKE<sup>73</sup>...

## Annex I

### Lep H1 construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPIMAN MFALILVIAT LVTGILWCVH MRLSERK<sup>67</sup>↓ETL  
GDV<sup>73</sup>...

### gp41 TM.5 construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPIMKLF IMIVGGLVGL RVRLSERKET LGDVTHR<sup>67</sup>↓ILT  
VPI<sup>73</sup>...

### VSV-G TM.5 (TM10) construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPMITSS IASFFFIHMHM RLSEKERTLG DVTHRIL<sup>67</sup>↓TVP  
IAQ<sup>73</sup>...

### VSV-G TM11 construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPMITSS IASFFFIIGH MRLSERKETL GDVTHRI<sup>67</sup>↓LTV  
PIA<sup>73</sup>...

### VSV-G TM14 construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPMITSS IASFFFIIGL IIMHRLSERK ETLGDVT<sup>67</sup>↓HRI  
LTV<sup>73</sup>...

### VSV-G TM17 construct

...PGCSSGQACE N<sup>1</sup>STGVTYSNV EPSDFVQTFS RRNGGEATSG  
FFEVPMITSS IASFFFIIGL IIGLFHMRLS ERKETLG<sup>67</sup>↓DVT  
HRI<sup>73</sup>...

### NAGK (-16)

...CEQSTGVTYS N<sup>1</sup>STSDVFQTF SRRNGGEATS GFFEVPMISR  
DDAAQVAKVL SEALPYIRRF VHMRLSERKE TLGDVTH<sup>67</sup>↓RIL...

### L9 (-16)

...CEQSTGVTYS N<sup>1</sup>STSDVFQTF SRRNGGEATS GFFEVPMIKA  
LEAQKQKEQR QAEEELANAK KHMRLSERKE TLGDVTH<sup>67</sup>↓RIL...

## ANNEX II

**Constructs analyzed in Chapter 2.** Amino acid sequences of the Lep G2 and G2' constructs. Glycosylation acceptor site NST or NVT are highlighted in bold at the appropriate positions. Cloned sequences between the GGPG flanking tetrapeptides are highlighted in yellow boxes. Charged residues in these yellow boxed regions are shown in red (Asp and Glu) or in blue (Lys and Arg). Amino acid substitutions with respect to the pHLIP3D are highlighted in bold.



**pHLIP3D only G1**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D N-7**

...ETKENGIRLS**NST**GGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D N-9**

...ETKENGIR**NST**TSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D N-11**

...ETKENG**NST**SETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D N-15**

...ET**NST**GIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D G2 or pHLIP3D C+5**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGG**NVT**  
 GQQQATWIVPPGQYFMMGDNRDNSA...

**pHLIP3D C+9**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
**QNV**TWIVPPGQYFMMGDNRDNSA...

**pHLIP3D C+11**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQ**QNV**TIVPPGQYFMMGDNRDNSA...

**pHLIP3D C+15**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATW**INVT**GQYFMMGDNRDNSA...

**pHLIP3D C+17**

...ETKENGIRLSETSGGPG**WARYADWLFTDLPLLLLDLLALLV**GPGGVPG  
 QQQATWIV**PNT**YFMMGDNRDNSA...

## Annex II

### **pHLIP3D G2' or pHLIP3D C+29**

...ETKENGIRLSETSGGPGWARYADWLFTDLPLLLLDLLALLVGPGGVPG  
QQQATWIVPPGQYFMMGDNRDNST...

### **Melittin G2**

...ETKENGIRLSETSGGPGIGAVLKVLTTGLPALISWIGPGGNVTGQQIAT  
WIVPPGQYFMMGDNRDNSA...

### **Melittin G2'**

...ETKENGIRLSETSGGPGIGAVLKVLTTGLPALISWIGPGGVPGQQIATW  
IVPPGQYFMMGDNRDNST...

### **pHLIP LL or pHLIP D11L G2**

...ETKENGIRLSETSGGPGWARYADWLFTLLPLLLLDLLALLVGPGGNVT  
GQQIATWIVPPGQYFMMGDNRDNSA...

### **pHLIP DD or pHLIP L12D G2**

...ETKENGIRLSETSGGPGWARYADWLFTDDPLLLLDLLALLVGPGGNVT  
GQQIATWIVPPGQYFMMGDNRDNSA...

### **pHLIP WW or pHLIP DL11/12WW G2**

...ETKENGIRLSETSGGPGWARYADWLFTWWPLLLLDLLALLVGPGGNVT  
GQQIATWIVPPGQYFMMGDNRDNSA...

### **pHLIP D18L**

...ETKENGIRLSETSGGPGWARYADWLFTDLPLLLLLLLALLVGPGGNVT  
GQQQATWIVPPGQYFMMGDNRDNSA...

### **pHLIP D18L**

...ETKENGIRLSETSGGPGWARYADWLFTDLPLLLDLDLLALLVGPGGNVT  
GQQQATWIVPPGQYFMMGDNRDNSA...

### **pHLIP LD17/18WW**

...ETKENGIRLSETSGGPGWARYADWLFTDLPLLLWLLALLVGPGGNVT  
GQQQATWIVPPGQYFMMGDNRDNSA...

### **pHLIP3D G3**

...ETKENGIRLSETSGGPGWARYADWLFTDLPLLLLDLLALLVGPGGNVT  
GQQQATWIVPPGQYFMMGDNRDNST...

**Melittin G3**

...ETKNGIRLSETSGGPGIGAVLKVLTTGLPALISWIGPGG**NVT**GQQIAT  
WIVPPGQYFMMGDNRD**NST**...





## ANNEX III

**Constructs analyzed in Chapter 2.** Amino acid sequences of sequences cloned in the Lep3G construct. Empty full length LepG3 sequence is shown first. Lep H1 and H2 are highlighted in gray boxes. G1, G2 and G3 glycosylation acceptor sites (NST or NVT) are highlighted in bold. Cloned sequences between the GGPG flanking tetrapeptide are highlighted in yellow. Charged residues in these yellow boxed regions are shown in red (Asp and Glu) or in blue (Lys and Arg). Amino acid substitutions with respect the WT pHLIP3D sequences are also highlighted in bold.



**LepG3 sequence**

MANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAAQAAAGDSLDKA  
 TLKKVAPKPGWLETGASVFPVLAIVLIVRSFIYEPFQIPSGSMPTLNS  
 TDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKR  
 AVGLPGDKVITYDPVSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQT  
 FSRNNGGEATSGFFEVPKQETKENGIRLSETSGGPGXXXGPGGNVTGQQ  
 QATWIVPPGQYFMMGDNRD**NST**DSRYWGFVPEANLVGRATAIWMSFDKQ  
 ESEWPTGLRLSRIGGIH-

**TCV HR1 or Insertion control**

GGPGKVLIVTGVLGLLLLIKWKGGPG

**pHLIP3D or pHLIP WT**

GGPGWARYADWLFTDLPLLLL DLLALLVGGPG

**#67 or Translocation control**

GGPGYNLYIYAFFASSLVHGFIGPGG

**pHLIP LL or pHLIP D11L**

GGPGWARYADWLFTLLPLLLL DLLALLVGGPG

**pHLIP DD pHLIP L12D**

GGPGWARYADWLFTDDPLLLL DLLALLVGGPG

**pHLIP WW or pHLIP DL11/12WW**

GGPGWARYADWLFTWWPLLLL DLLALLVGGPG

**pHLIP AA**

GGPGWARYADWLFTAAPLLL DLLALLVGGPG

**pHLIP II**

GGPGWARYADWLFTIIPLLL DLLALLVGGPG

**pHLIP KK**

GGPGWARYADWLFTKKPLLLL DLLALLVGGPG

**pHLIP AAA**

GGPGWARYADWLFTAAALLL DLLALLVGGPG

Annex III

**pHLIP CCC**

GGPGWARYADWLFT**CCCLLLL**DLLALLVGPGG

**pHLIP DDD**

GGPGWARYADWLFT**DDDLLLL**DLLALLVGPGG

**pHLIP EEE**

GGPGWARYADWLFT**EEELLLL**DLLALLVGPGG

**pHLIP FFF**

GGPGWARYADWLFT**FFFLLLL**DLLALLVGPGG

**pHLIP GGG**

GGPGWARYADWLFT**GGGLLLL**DLLALLVGPGG

**pHLIP HHH**

GGPGWARYADWLFT**HHHLLL**DLLALLVGPGG

**pHLIP III**

GGPGWARYADWLFT**IIILLLL**DLLALLVGPGG

**pHLIP KKK**

GGPGWARYADWLFT**KKKLLL**DLLALLVGPGG

**pHLIP LLL**

GGPGWARYADWLFT**LLLLLLL**DLLALLVGPGG

**pHLIP MMM**

GGPGWARYADWLFT**MMMLLLL**DLLALLVGPGG

**pHLIP NNN**

GGPGWARYADWLFT**NNNLLL**DLLALLVGPGG

**pHLIP PPP**

GGPGWARYADWLFT**PPPLLLL**DLLALLVGPGG

**pHLIP QQQ**

GGPGWARYADWLFT**QQQLLLL**DLLALLVGPGG

**pHLIP RRR**

GGPGWARYADWLFT**RRR**LLLLDLLALLVGP

**pHLIP SSS**

GGPGWARYADWLFT**SSS**LLLLDLLALLVGP

**pHLIP TTT**

GGPGWARYADWLFT**TTT**LLLLDLLALLVGP

**pHLIP VVV**

GGPGWARYADWLFT**VVV**LLLLDLLALLVGP

**pHLIP WWW**

GGPGWARYADWLFT**WWW**LLLLDLLALLVGP

**pHLIP YYY**

GGPGWARYADWLFT**YYY**LLLLDLLALLVGP

**pHLIP GAA**

GGPGWARYADWLFT**GAA**LLLLDLLALLVGP

**pHLIP AGA**

GGPGWARYADWLFT**AGA**LLLLDLLALLVGP

**pHLIP AAG**

GGPGWARYADWLFT**AAG**LLLLDLLALLVGP

**pHLIP GGA**

GGPGWARYADWLFT**GGA**LLLLDLLALLVGP

**pHLIP GAG**

GGPGWARYADWLFT**GAG**LLLLDLLALLVGP

**pHLIP AGG**

GGPGWARYADWLFT**AGG**LLLLDLLALLVGP

**2L/17A**

GGPGAAAAAAAA**L**AAAA**L**AAAAAAAAGP

**3L/16A**

GGPGAAAAAAAA**L**L**L**AAAAAAAAGP

Annex III

**4L/15A v1**

GGPGAAALAAALAAAAALAAALAAAGPGG

**4L/15A v2**

GGPGAAAAALALAAAAALALAAAAAGPGG

**5L/14A**

GGPGAAALAAALAAALAAALAAALAAAGPGG

## **ANNEX IV**

**Supplemental calculations of amino acid preference for inserted state analyzed in Chapter 2.**





$$p_{\text{inserted}} = f_{1g} / (f_{1g} + f_{2g} + f_{3g})$$

$$K_{\text{app}}^{I/S} = K_{\text{app}}^{S/I^{-1}} = f_{1g} / f_{2g}; \quad K_{\text{app}}^{I/T} = K_{\text{app}}^{T/I^{-1}} = f_{1g} / f_{3g}$$

$$\Delta G_{\text{app}}^{I/S} = -\Delta G_{\text{app}}^{S/I} = -RT \ln K_{\text{app}}^{I/S}$$

$$\Delta G_{\text{app}}^{I/T} = -\Delta G_{\text{app}}^{T/I} = -RT \ln K_{\text{app}}^{I/T}$$

$$K_{\text{app}}^{\text{inserted}} = K_{\text{app}}^{I/S} \cdot K_{\text{app}}^{I/T} = f_{1g} / f_{2g} \cdot f_{1g} / f_{3g}$$

$$\Delta G_{\text{app}}^{\text{inserted}} = -RT \ln K_{\text{app}}^{\text{inserted}}$$



## **ANNEX V**

**Supplemental calculations of amino acid preference for translocated state analyzed in Chapter 2.**



$$p_{translocated} = f_{3g} / (f_{1g} + f_{2g} + f_{3g})$$

$$K_{app}^{T/I} = K_{app}^{I/T^{-1}} = f_{3g} / f_{1g}; \quad K_{app}^{T/S} = K_{app}^{S/T^{-1}} = f_{3g} / f_{2g}$$

$$\Delta G_{app}^{T/I} = -\Delta G_{app}^{I/T} = -RT \ln K_{app}^{T/I}$$

$$\Delta G_{app}^{T/S} = -\Delta G_{app}^{S/T} = -RT \ln K_{app}^{T/S}$$

$$K_{app}^{translocated} = K_{app}^{T/I} \cdot K_{app}^{T/S} = f_{3g} / f_{1g} \cdot f_{3g} / f_{2g}$$

$$\Delta G_{app}^{translocated} = -RT \ln K_{app}^{translocated}$$



## ANNEX VI

**Bacteriorhodopsin sequence.** Amino acid sequences cloned in pGEM. Predicted TMDs are highlighted in gray boxes. The last residue for the truncated variants is indicated in with an arrow ( $\blacktriangleleft$ ). Gln 105 residue is highlighted in bold. Inserted *bbe* loop is highlighted in a yellow box. The glycosylation acceptor site in the *bbe* sequence is underlined.





**Bacteriorhodopsin full sequence (without SS)**

MQAQITGRPEWVWLALGTALMGLGTLYFLVKGMGVSDPDAK<sup>41</sup>↓FYAI  
 TTLVPAIAFTMYLSMLLGYGLTMVFPFGGEQNPI<sup>78</sup>↓YWARYADWLFTTP  
 LLLLDLALLVDADQGT<sup>107</sup>↓ILALVGADGIMIGTGLVGALTKVYSYR<sup>134</sup>  
 ↓FVWWAISTAAMLYILYVLFFGFTSKAESMRPEVA<sup>168</sup>↓STFKVLRNVT  
 VVLWSAYPVVWLI GSEGAGIVPLNIET<sup>205</sup>↓LLFMVLDVSAKVGFGLLI  
 LRSRAIFGEAEAPEPSAGDGAAATS-

**Bacteriorhodopsin with an inserted loop from bbe**

MQAQITGRPEWVWLALGTALMGLGTLYFLVKGMGVSDPDAK<sup>41</sup>FYAITTL  
 VPAIAFTMYLSMLLGYGLTMVFPFGGEQNPIYWARYADWLFTTP LLLLDL  
 ALLVDADTSMVNDNLLSSCLNSHG<sup>VHNFTTLSTDTNSDYFKLLHASMQN</sup>  
<sup>PLFAKPTVSKPSFIVMPGSTS</sup>QGT<sup>107</sup>↓ILALVGADGIMIGTGLVGAL  
 T KVYSYR<sup>134</sup>↓FVWWAISTAAMLYILYVLFFGFTSKAESMRPEVASTFKV  
 LRNVT VVLWSAYPVVWLI GSEGAGIVPLNIETLLFMVLDVSAKVGFGLI  
 LRSRAIFGEAEAPEPSAGDGAAATS-



## **PUBLICATIONS DURING THE THESIS**





# Membrane insertion and topology of the translocon-associated protein (TRAP) gamma subunit



Manuel Bañó-Polo, Carlos A. Martínez-Garay, Brayan Grau, Luis Martínez-Gil, Ismael Mingarro \*

Departament de Bioquímica i Biologia Molecular, Estructura de Recerca Interdisciplinària en Biotecnologia i Biomedicina (ERI BioTecMed), Universitat de València, E-46 100 Burjassot, Spain

## ARTICLE INFO

### Article history:

Received 17 November 2016  
Received in revised form 19 January 2017  
Accepted 25 January 2017  
Available online 26 January 2017

### Keywords:

Endoplasmic reticulum  
Membrane protein topology  
N-linked glycosylation  
Sec61 complex  
Translocon-associated proteins

## ABSTRACT

Translocon-associated protein (TRAP) complex is intimately associated with the ER translocon for the insertion or translocation of newly synthesised proteins in eukaryotic cells. The TRAP complex is comprised of three single-spanning and one multiple-spanning subunits. We have investigated the membrane insertion and topology of the multiple-spanning TRAP- $\gamma$  subunit by glycosylation mapping and green fluorescent protein fusions both *in vitro* and in cell cultures. Results demonstrate that TRAP- $\gamma$  has four transmembrane (TM) segments, an Nt/Ct cytosolic orientation and that the less hydrophobic TM segment inserts efficiently into the membrane only in the cellular context of full-length protein.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

In eukaryotic cells insertion of most integral membrane proteins into the endoplasmic reticulum (ER) membrane occurs primarily in a co-translational manner. In this process, targeting of the ribosome–mRNA–nascent chain complex to the ER depends on the signal recognition particle (SRP) and its interaction with the membrane-bound SRP receptor. The ribosome and the nascent chain are then transferred to the translocon, a multi-protein complex that facilitates insertion of integral membrane proteins into the lipid bilayer and translocation of secreted proteins across the lipid bilayer [1]. Translocons are not passive pores in the bilayer, but instead are dynamic complexes that cycle between ribosome-bound and ribosome-free states, and convert between translocation and membrane integration modes of action, while maintaining the membrane permeability barrier [2,3]. The core components of the mammalian translocon are the Sec61  $\alpha$ ,  $\beta$  and  $\gamma$  subunits [4] and the translocating chain-associating membrane protein [5,6]. However, several other proteins including the signal peptidase, the oligosaccharyltransferase (OST), PAT-10, RAMP4, BAP31 or the translocon-associated protein (TRAP) complex, interact at some point with the core translocon modulating its activity [1].

The TRAP complex is comprised of four membrane protein subunits. The  $\alpha$ ,  $\beta$ , and  $\delta$ -subunits are single-spanning proteins with suggested large N-terminal (Nt) luminal and small C-terminal (Ct) cytosolic domains, while  $\gamma$ -subunit was proposed to cross the membrane four times with a putative Nt/Ct luminal orientation [7]. This complex remains stably associated with detergent-solubilised ribosome-translocon complexes [8], being an integral part of the translocon. Despite notorious efforts, the role of TRAP complex during membrane insertion or translocation is not fully understood. It has been proposed that TRAP complex acts, in a substrate-specific manner, facilitating the initiation of protein translocation [9]. It has also been suggested a role of TRAP complex in membrane topology regulation by moderating the 'positive-inside' rule [10] of membrane proteins with weak topogenic determinants [11]. The TRAP complex is also involved in ER associated degradation, where its function might be to recruit misfolded proteins to the translocon [12]. Most of these studies have been carried out using *in vitro* reconstituted systems, mainly with microsomal membranes. Its role in membrane protein biogenesis has been emphasised *in vivo*, where it has been shown that TRAP- $\alpha$  subunit mutant alters mammalian heart development [13], a mutation in TRAP- $\delta$  has been related to human congenital disorders of glycosylation [14], and TRAP- $\gamma$  subunit plays an important role in the process of pronephros differentiation during *Xenopus* development [15], and appears to be required for vascular network formation in murine placental development [16]. Despite the general importance of the TRAP complex in secreted and membrane protein biogenesis, and the particular involvement of TRAP- $\gamma$  subunit in development, its membrane topology has not been thoroughly investigated.

**Abbreviations:** ER, endoplasmic reticulum; GFP, green fluorescent protein; HR, hydrophobic region; Lep, leader peptidase; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis; SRP, signal recognition particle; TM, transmembrane; TRAP, translocon-associated protein.

\* Corresponding author.

E-mail address: [Ismael.Mingarro@uv.es](mailto:Ismael.Mingarro@uv.es) (I. Mingarro).



# The role of hydrophobic matching on transmembrane helix packing in cells

Brayan Grau<sup>1</sup>, Matti Javanainen<sup>2,3</sup>, María Jesús García-Murria<sup>1</sup>, Waldemar Kulig<sup>2,3</sup>, Ilpo Vattulainen<sup>2,3,4</sup>, Ismael Mingarro<sup>1</sup>, Luis Martínez-Gil<sup>1,\*</sup>

<sup>1</sup> Departamento de Bioquímica y Biología Molecular, ERI BioTecMed, Universitat de València, E-46100 Burjassot, Spain.

<sup>2</sup> Laboratory of Physics, Tampere University of Technology, FI-33101 Tampere, Finland.

<sup>3</sup> Department of Physics, University of Helsinki, POB 64, FI-00014 Helsinki, Finland.

<sup>4</sup> MEMPHYS - Centre for Biomembrane Physics.

\* Corresponding Author:

Luis Martínez-Gil, Departamento de Bioquímica y Biología Molecular, ERI BioTecMed, Universitat de València, E-46100 Burjassot, Spain; E-mail: luis.martinez-gil@uv.es

**ABSTRACT** Folding and packing of membrane proteins are highly influenced by the lipidic component of the membrane. Here, we explore how the hydrophobic mismatch (the difference between the hydrophobic span of a transmembrane protein region and the hydrophobic thickness of the lipid membrane around the protein) influences transmembrane helix packing in a cellular environment. Using a ToxRED assay in *Escherichia coli* and a Bimolecular Fluorescent Complementation approach in human-derived cells complemented by atomistic molecular dynamics simulations we analyzed the dimerization of Glycophorin A derived transmembrane segments. We concluded that, biological membranes can accommodate transmembrane homo-dimers with a wide range of hydrophobic lengths. Hydrophobic mismatch and its effects on dimerization are found to be considerably weaker than those previously observed in model membranes, or under *in vitro* conditions, indicating that biological membranes (particularly eukaryotic membranes) can adapt to structural deformations through compensatory mechanisms that emerge from their complex structure and composition to alleviate membrane stress. Results based on atomistic simulations support this view, as they revealed that Glycophorin A dimers remain stable, despite of poor hydrophobic match, using mechanisms based on dimer tilting or local membrane thickness perturbations. Furthermore, hetero-dimers with large length disparity between their monomers are also tolerated in cells, and the conclusions that one can draw are essentially similar to those found with homo-dimers. However, large differences between transmembrane helices length hinder the monomer/dimer equilibrium, confirming that, the hydrophobic mismatch has, nonetheless, biologically relevant effects on helix packing *in vivo*.

doi: 10.15698/cst2017.11.111  
Received originally: 10.05.2017;  
in revised form: 13.10.2017,  
Accepted 16.10.2017,  
Published 02.11.2017.

**Keywords:** hydrophobic match, mismatch, Glycophorin A, membrane protein folding, helix packing, transmembrane domain dimerization.

**Abbreviations:**  
BIFC - Bimolecular Fluorescent Complementation,  
ER - endoplasmic reticulum,  
GpA - Glycophorin A,  
MBP - maltose binding protein,  
RFP - red fluorescent protein,  
TM - transmembrane,  
VFP - venus fluorescent protein.

## INTRODUCTION

Assembly of the native structure of most integral membrane proteins takes place in two main steps [1]. The first step includes targeting and insertion of the protein into a lipid membrane. In the case of alpha-helical membrane proteins this initial step occurs generally co-translationally (coupled with the translation of the protein) through the

translocon, a multiprotein complex that facilitates not only the insertion of integral membrane proteins into the lipid bilayer but also translocation of soluble proteins into the endoplasmic reticulum (ER) lumen [2]. In the second stage, if required, the transmembrane (TM) segments interact to form the tertiary and quaternary structure of the mature functional membrane protein.





ARTICLE

DOI: 10.1038/s41467-018-07554-7

OPEN

# Transmembrane but not soluble helices fold inside the ribosome tunnel

Manuel Bañó-Polo<sup>1</sup>, Carlos Baeza-Delgado<sup>1</sup>, Silvia Tamborero<sup>1</sup>, Anthony Hazel<sup>2</sup>, Brayan Grau<sup>1</sup>, IngMarie Nilsson<sup>3</sup>, Paul Whitley<sup>4</sup>, James C. Gumbart<sup>2</sup>, Gunnar von Heijne<sup>3</sup> & Ismael Mingarro<sup>1</sup>

Integral membrane proteins are assembled into the ER membrane via a continuous ribosome-translocon channel. The hydrophobicity and thickness of the core of the membrane bilayer leads to the expectation that transmembrane (TM) segments minimize the cost of harbouring polar polypeptide backbones by adopting a regular pattern of hydrogen bonds to form  $\alpha$ -helices before integration. Co-translational folding of nascent chains into an  $\alpha$ -helical conformation in the ribosomal tunnel has been demonstrated previously, but the features governing this folding are not well understood. In particular, little is known about what features influence the propensity to acquire  $\alpha$ -helical structure in the ribosome. Using in vitro translation of truncated nascent chains trapped within the ribosome tunnel and molecular dynamics simulations, we show that folding in the ribosome is attained for TM helices but not for soluble helices, presumably facilitating SRP (signal recognition particle) recognition and/or a favourable conformation for membrane integration upon translocon entry.

<sup>1</sup>Estructura de Recerca Interdisciplinària en Biotecnologia i Biomedicina (ERI BioTecMed), Departament de Bioquímica i Biologia Molecular, Universitat de València, E-46100 Burjassot, Spain. <sup>2</sup>School of Physics, School of Chemistry and Biochemistry, Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA. <sup>3</sup>Center for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm, Sweden. <sup>4</sup>Department of Biology and Biochemistry, Centre for Regenerative Medicine, University of Bath, Bath BA2 7AY, UK. These authors contributed equally: Manuel Bañó-Polo, Carlos Baeza-Delgado, Silvia Tamborero, Anthony Hazel, Brayan Grau. Correspondence and requests for materials should be addressed to I.M. (email: [Ismael.Mingarro@uv.es](mailto:Ismael.Mingarro@uv.es))



## Controllable membrane remodeling by a modified fragment of the apoptotic protein Bax

View Article Online  
DOI: 10.1039/D0FD00070A

Katherine G. Schaefer<sup>1,\*</sup>, Brayán Grau<sup>2,3,\*</sup>, Nicolas Moore<sup>2</sup>, Ismael Mingarro<sup>3</sup>, Gavin M. King<sup>1,4</sup> and Francisco N. Barrera<sup>2</sup>

<sup>1</sup> Department of Physics and Astronomy, University of Missouri, Columbia, MO 65211, USA

<sup>2</sup> Department of Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville, 37996, USA.

<sup>3</sup> Departament de Bioquímica i Biologia Molecular, Estructura de Recerca Interdisciplinària en Biotecnologia i Biomedicina (ERI BioTecMed), Universitat de València. E-46100 Burjassot, Spain.

<sup>4</sup> Department of Biochemistry, University of Missouri, Columbia, Missouri, 65211, USA.

\* These authors contributed equally

Correspondence: [fbarrera@utk.edu](mailto:fbarrera@utk.edu), [kinggm@missouri.edu](mailto:kinggm@missouri.edu)





**Cite this article:** Duart G, García-Murria MJ, Grau B, Acosta-Cáceres JM, Martínez-Gil L, Mingarro I. 2020 SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol.* **10**: 200209.

<http://dx.doi.org/10.1098/rsob.200209>

Received: 10 July 2020

Accepted: 18 August 2020

**Subject Area:**

biochemistry/molecular biology

**Keywords:**

coronavirus, envelope protein, membrane insertion, SARS-CoV-2, topology

**Authors for correspondence:**

Luis Martínez-Gil

e-mail: [luis.martinez-gil@uv.es](mailto:luis.martinez-gil@uv.es)

Ismael Mingarro

e-mail: [ismael.mingarro@uv.es](mailto:ismael.mingarro@uv.es)

<sup>†</sup>Equal contribution (alphabetical order).

# SARS-CoV-2 envelope protein topology in eukaryotic membranes

Gerard Duart<sup>†</sup>, M<sup>a</sup> Jesús García-Murria<sup>†</sup>, Brayan Grau<sup>†</sup>, José M. Acosta-Cáceres, Luis Martínez-Gil and Ismael Mingarro

Departament de Bioquímica i Biologia Molecular, Estructura de Recerca Interdisciplinar en Biotecnologia i Biomedicina (ERI BioTecMed), Universitat de València E-46100 Burjassot, Spain

**id** MJG-M, 0000-0002-8396-0335; BG, 0000-0002-8582-023X; LM-G, 0000-0002-9076-7760; IM, 0000-0002-1910-1229

Coronavirus E protein is a small membrane protein found in the virus envelope. Different coronavirus E proteins share striking biochemical and functional similarities, but sequence conservation is limited. In this report, we studied the E protein topology from the new SARS-CoV-2 virus both in microsomal membranes and in mammalian cells. Experimental data reveal that E protein is a single-spanning membrane protein with the N-terminus being translocated across the membrane, while the C-terminus is exposed to the cytoplasmic side (N<sub>t</sub><sub>lum</sub>/C<sub>t</sub><sub>cyt</sub>). The defined membrane protein topology of SARS-CoV-2 E protein may provide a useful framework to understand its interaction with other viral and host components and contribute to establish the basis to tackle the pathogenesis of SARS-CoV-2.

## 1. Introduction

Coronavirus disease 2019 (COVID-19), an extremely infectious human disease caused by coronavirus SARS-CoV-2, has spread around the world at an unprecedented rate, causing a worldwide pandemic. While the number of confirmed cases continues to grow rapidly, the molecular mechanisms behind the biogenesis of viral proteins are not fully unravelled. The SARS-CoV-2 genome encodes up to 29 proteins, although some may not get expressed [1]. The viral RNA is packaged by the structural proteins to assemble viral particles at the ERGIC (ER-Golgi intermediate compartment). The four major structural proteins are the spike (S) surface glycoprotein, the membrane (M) matrix protein, the nucleocapsid (N) protein, and the envelope (E) protein. These conserved structural proteins are synthesized from sub-genomic RNAs (sgRNA) encoded close to the 3' end of the viral genome [2].

Among the four major structural proteins, the E protein is the smallest and has the lowest copy number of the membrane proteins found in the lipid envelope of mature virus particles (reviewed [3,4]). However, it is critical for pathogenesis of other human coronaviruses [5,6]. Interestingly, the sgRNA encoding E protein is one of the most abundantly expressed transcripts despite the protein having a low copy number in mature viruses [1]. It encodes a 75 residues long polypeptide with a predicted molecular weight of approximately 8 kDa. Two aliphatic amino acids (Leu and Val) constitute a substantial portion (36%, 27/75) of the E protein, which accounts for the high grand average of hydrophobicity (GRAVY) index of the protein (1.128), as calculated using the ExPASy ProtParam tool (<https://web.expasy.org/protparam/>). Comparative sequence analysis of the E protein of SARS-CoV-2 and the other six known human coronaviruses do not reveal any large homologous/identical regions (figure 1), with only the initial methionine, Leu39, Cys40 and Pro54 being ubiquitously conserved. With regard to overall sequence similarity SARS-CoV-2 E protein has the highest similarity to SARS-CoV (94.74%) with only minor differences (figure 1b), followed by MERS-CoV (36.00%). Interestingly, sequence similarities are significantly lower for the other four human coronaviruses, which usually cause mild to moderate



ARTICLE



<https://doi.org/10.1038/s41467-020-19881-9>

OPEN

# Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control cellular apoptosis

Maria Jesús García-Murria<sup>1,3</sup>, Gerard Duart<sup>1,3</sup>, Brayan Grau<sup>1</sup>, Elisabet Diaz-Beneitez<sup>2</sup>, Dolores Rodríguez<sup>2</sup>, Ismael Mingarro<sup>1</sup> & Luis Martínez-Gil<sup>1</sup>✉

Viral control of programmed cell death relies in part on the expression of viral analogs of the B-cell lymphoma 2 (Bcl2) protein known as viral Bcl2s (vBcl2s). vBcl2s control apoptosis by interacting with host pro- and anti-apoptotic members of the Bcl2 family. Here, we show that the carboxyl-terminal hydrophobic region of herpesviral and poxviral vBcl2s can operate as transmembrane domains (TMDs) and participate in their homo-oligomerization. Additionally, we show that the viral TMDs mediate interactions with cellular pro- and anti-apoptotic Bcl2 TMDs within the membrane. Furthermore, these intra-membrane interactions among viral and cellular proteins are necessary to control cell death upon an apoptotic stimulus. Therefore, their inhibition represents a new potential therapy against viral infections, which are characterized by short- and long-term deregulation of programmed cell death.

<sup>1</sup>Department of Biochemistry and Molecular Biology, Institut de Biociències i Biomedicina, Universitat de València, 46100 Burjassot, Spain. <sup>2</sup>Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma, 28049 Madrid, Spain. <sup>3</sup>These authors contributed equally: Maria Jesús García-Murria, Gerard Duart. ✉email: [Luis.martinez-gil@uv.es](mailto:Luis.martinez-gil@uv.es)