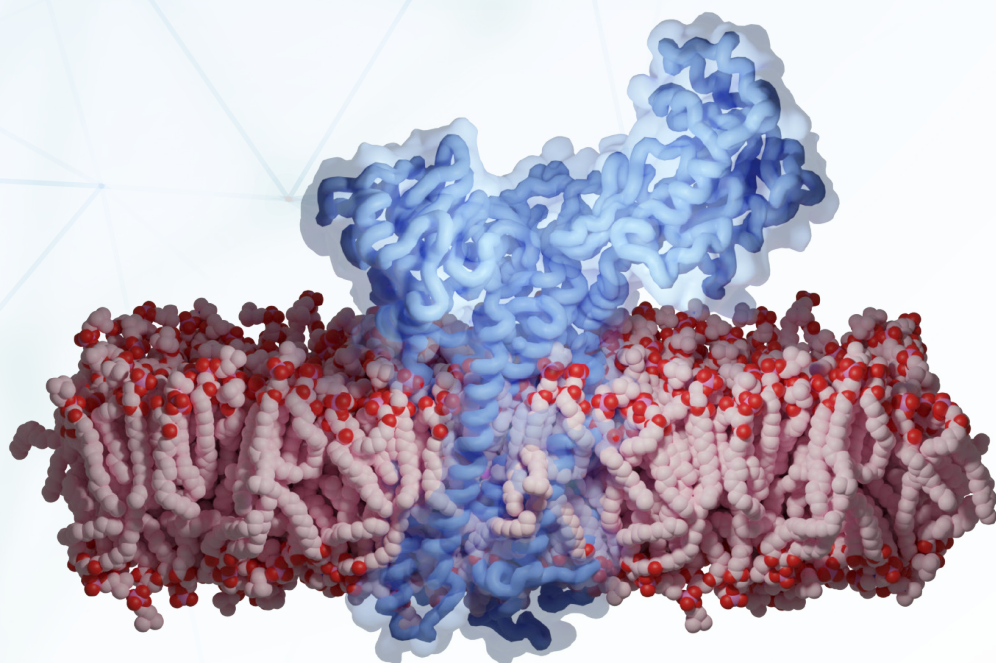# Insertion and interaction of transmembrane helices, from basic principles to rational design

Doctoral Programme in Biomedicine and Biotechnology

Gerard Duart i Celdrán

Directed by:
Dr. Ismael Mingarro
Dr. Luis Martínez-Gil

VNIVERSITAT DE VALÈNCIA

September 2022

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 Contractat Doctor del Departament de Bioquímica i Biologia Molecular de la Universitat de València.

MANIFESTEN que, En GERARD DUART CELDRÁN, Graduat en Biotecnologia per la Universitat de València, ha realitzat sota la nostra direcció al Laboratori de Proteïnes de Membrana de l'Institut de Biotecnologia i Biomedicina (BIOTECMED) i finançat amb un Contracte del Programa de "Formación de Profesorado Universitario (FPU18/05771) del Ministerio de Educacion" el treball que amb el títol: "*Insertion and interaction of transmembrane segments, from basic principles to rational design*", presenta per a optar al grau de Doctor per la Universitat de València. Part d'aquest treball el va realitzar al *"Weizmann Institute of Science"* (Israel) en una estada finançada pel Vicerectorat de Relacions Internacionals de 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 el 15 de setembre de 2022.

Dr. Ismael Mingarro   Dr. Luis Martínez-Gil

Science is not about building a body of known "facts". It is a method for asking awkward questions and subjecting them to a reality-check, thus avoiding the human tendency to believe whatever makes us feel good.

-Terry Pratchett

Abstract

Membrane proteins are an important group of macromolecules that play central roles in several vital cellular processes. Although they represent around the 30% of the genes in most genomes, the information that we have about them is far away from their biological abundance and relevance. The main goal of this thesis is to delve in our understanding of membrane protein biogenesis. We studied α-helical membrane protein segments, from basic principles to rational design using some of the more relevant biochemical techniques. This work led us to a better description of how polar residues can be inserted into the membrane, improving our understanding of the insertion process and the membrane protein topology determination. We also explored the complexity of transmembrane-transmembrane interactions and their role to fine-tune apoptotic networks. Finally, we used these interactions to computationally design transmembrane inhibitors as new possible therapeutic agents to regulate cellular death.

Table of contents

Abbreviations

Bcl2: B-cell lymphoma 2
BiFC: Bimolecular Fluorescence Complementation
BP: Bis(monoacylglycero)phosphate
cBcl2: Cellular Bcl2
Chl: Cholesterol
CL: Cardiolipin
Cryo-EM: Cryogenic electron microscopy
Cryo-ET: Cryogenic electron tromography
Ct: C-terminal; Carboxy-terminal
DOPC: Dioleoylphosphatidylcholine
E protein: Envelope protein
eGFP: Enhanced green fluorescent protein
EMC: Endoplasmatic reticulum membrane protein complex
ER: Endoplasmatic reticulum
FL: Full length
GET: Guided entry of tail-anchored
GpA: Glycophorin A
IMP: Integral membrane protein
Lep: Leader peptidase
Lo: liquid-ordered
MERS-CoV: Middle East respiratory syndrome coronavirus
NAC: Nascent polypeptide-associated complexes
Nt: N-terminal; Amino-terminal
OST: Oligosaccharyltransferase
PAT10: Protein associated with the translocon of 10 kDa

PC: Phosphatidylcholine
PDB: Protein Data Bank
PE: Phosphatidylethanolamine
PG: Phosphatidylglycerol
PI: Phosphatidylinositol
PMP: Peripheral membrane protein
PPI: Protein-protein interactions
PS: Phosphatidylserine
RM: Rough microsomes
SARS-CoV: Severe acute respiratory syndrome coronavirus
SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2
SGTA: Small Glutamine rich Tetratricopeptide repeat co-chaperone Alpha
SM: Sphingomyelin
SPA: Single particle analysis
SRP: Signal recognition particle
SWATH-MS: Sequential window acquisition of all theoretical mass spectra
T20: TOMM20
TA: Tail anchored
TM: Transmembrane
TMD: Transmembrane domain
TMHOP: Trans-membrane Homo Oligomer Predictor
TRAM: Translocation associated membrane protein
vBcl2: Viral Bcl2
VFP: Venus fluorescent protein
wt: Wild type

Amino acids letter code

Alanine – Ala – A

Arginine – Arg – R

Asparagine – Asn – N

Aspartic acid – Asp – D

Cysteine – Cys – C

Glutamine – Gln – Q

Glutamic acid – Glu – E

Glycine – Gly – G

Histidine – His – H

Isoleucine – Ile – I

Leucine – Leu – L

Lysine – Lys – K

Methionine – Met – M

Phenilalanine – Phe – F

Proline – Pro – P

Serine – Ser – S

Threonine – Thr – T

Tryptophan – Trp – W

Tyrosine – Tyr – Y

Valine – Val – V

1. Introduction

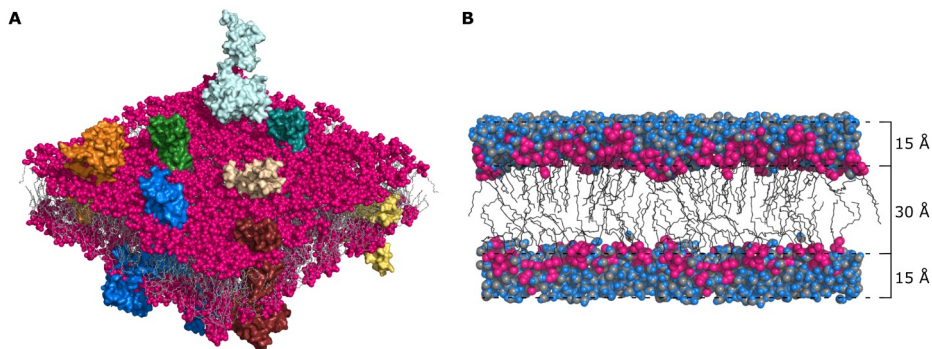## 1.1 Biological membranes and membrane proteins

The biological membrane is a physical barrier composed mainly by lipids and proteins, not only between the cell content and the environment but also between the cell compartments and the cytoplasm in eukaryotic cells. Apart from being a barrier, the membrane is also the only cellular surface of interaction with the external media. Consequently, it plays a central role in the communication and signalling pathways.

Our idea about the biological membrane organization is strongly influenced by the classical fluid mosaic membrane model proposed by Singer and Nicolson in 1972 (Singer and Nicolson, 1972). This early model combined the idea of the membrane being a bidimensional lipid-based fluid structure with the presence of embedded proteins that are free to diffuse laterally in a mosaic manner. This theory supposed a change of paradigm contrasting with the previous tri-layer membrane model based on the lipid bilayer proposal of Gorter and Grendel almost a century ago (Gorter and Grendel, 1925). After 50 years, the fluid mosaic model remains still relevant for depicting the basic organization of a huge diversity of intracellular and cellular membranes from lower forms of life to plant and animal cells (Nicolson, 2014). Initially, as it was proposed, the fluid mosaic model described the biological membrane as a matrix made up of mostly phospholipids with mobile globular integral membrane proteins and glycoproteins that were intercalated into the bilayer (Figure 1A). The confirmation of this model came over the years with numerous studies on the structure of membrane phospholipids and their lateral motion in the membrane plane (Edidin, 2003; Sanderson, 2012). Although the fluid mosaic membrane model is considered the most successful general model of biological membranes, as many other scientific models, it has been modified from its original proposal to reflect new observations that were not anticipated when initially proposed. Some examples of these new incorporations are the presence of lipid specific domains, as lipid rafts, and the interactions between membrane-associated cytoskeletal components and

their importance on distributing trans-membrane glycoproteins. In these cases, both suggest less membrane fluidity and dynamism than initially proposed (Israelachvili, 1977; Nicolson, 2014; Vereb et al., 2003).

The membrane is composed by small amphiphilic molecules which incorporate a hydrophilic head group and a hydrophobic tail group. These are mainly phospholipids but also include other lipids (like sphingolipids and sterols) (Harayama and Riezman, 2018). These amphipathic molecules self-associate through their hydrophobic tails excluding water to form bilayers driven by van der Waals forces and the hydrophobic effect (Singer, 1974). Based on their physical chemical properties we can distinguish two major regions in a lipid bilayer: the hydrophobic inner core formed mainly by the hydrocarbon tails of the lipids and the hydrophilic region populated with the polar head groups. The first measurements about the dimensions of these regions were performed by Wiener and White in 1992 using dioleoylphosphatidylcholine (DOPC) as a model lipid and concluded that the hydrophobic core of the membrane spreads around 30 Å, contributing in 15 Å each side of the bilayer, whereas the polar region was also around 15 Å (Figure 1B) (Stillwell, 2013; Wiener and White, 1992).

Figure 1. Biochemical structure of a biological membrane. (A) Schematic representation of the fluid mosaic membrane model proposed by Singer and Nicolson in 1972. The proteins can be found floating in the lipid bilayer partially or totally inserted. **(B)** Simulated model membrane of DOPC. The dimensions in Å for each region are depicted. The hydrophobic core expands around 30 Å while the polar region (interfacial region + polar head group region) fills around 15 Å.

It is important to point out that in biological membranes there are differences between the composition of the inner and the outer monolayers. This asymmetric distribution of the membrane components in the bilayer was known for some time before the fluid mosaic model was published (Nicolson, 2014; Stoeckenius and Engelman, 1969). For example, phospholipids, sugars, and proteins are asymmetrically distributed between the inner and the outer membrane leaflets. Cells must maintain this asymmetric distribution of outer and inner membrane proteins (enzymes, receptors, etc.) and other membrane components for the correct operation of every organism. For this purpose, the lipid asymmetry in cell membranes is crucial for facilitating the appropriate display of receptors, adhesion molecules, signalling systems, scaffolding structures etc., but also for guiding membrane curvature and structure (Nicolson, 2014). The asymmetric incorporation of proteins, sugars and other molecular species within the phospholipids results in the production of the fully functioning cell membrane that can control the composition of the intracellular environment through tightly regulated molecular machineries (Hilton et al., 2021).

Although biological membranes share similarities in all living organism, some peculiarities deserve to be mentioned. Bacterial plasma membranes are composed by proteins and lipids in similar proportions and organized in a classic bilayer structure. In terms of lipid composition, bacterial membranes present a large diversity. Different bacterial species display different membrane compositions and even the membrane composition of cells belonging to a single species is not constant depending on the environment (Sohlenkamp and Geiger, 2016). Generally, the outer membrane in bacteria contain mixtures of polar phospholipids such as phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL). Traditionally, the bacterial membrane has been modelled in synthetic systems using a mixture of PE:PG in a 3:1 ratio. This conventional mixture is very similar to the membrane composition of some Gram-negative bacteria such as *Escherichia coli* (PE:PG:CL 75:20:5), one of the most commonly used model organisms in laboratories worldwide (Hilton et al., 2021; Murzyn et al., 2005).

In eukaryotic cells, membrane composition differs a lot depending on the organism, cell type, membrane leaflets and membrane subdomains. As mammalian cells are compartmented cells, we can account for different types of membranes: the cytoplasmic membrane and the membranes of the organelles (Harayama and Riezman, 2018). In these membranes, lipids are distributed heterogeneously in different ranges: subcellular organelles show varied lipid arrangements that can be hardly standardized. Lipid proportions in the different organelle membranes are summarized in Table 1 (Hullin-Matsuda et al., 2014). Interestingly, the mitochondrial membrane is the only one in eukaryotic cells containing a significant amount of CL and no cholesterol in it, likely derived from its prokaryotic origin (van Meer et al., 2008). This composition complexity is key to determine the specific functions and identities of the different organelles. The lipid composition affects the membrane physical properties such as membrane rigidity and thus the function of the proteins inserted in it (Harayama and Riezman, 2018). The secretory pathway is the best characterized model of lipid distribution, and it involves the endoplasmatic reticulum (ER), the Golgi apparatus, and the plasmatic membrane. These three compartments display growing amounts of cholesterol and sphingolipids concentrations, with higher levels in the plasma membrane. In fact, this cholesterol and sphingolipids gradient is responsible of the increased membrane thickness and the stiffness of the plasma membrane (Casares et al., 2019; Hullin-Matsuda et al., 2014).

Table 1. Lipid composition of cellular eukaryotic membranes. Cholesterol (Chl), phosphatidylcholine (PC), sphingomyelin (SM), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI), bis(monoacylglycero)phosphate (BP) and cardiolipin (CL) proportions in the membranes of the different organelles.

Membrane	Chl	PC	SM	PE	PS	PI	BP	CL
Mitochondria	-	37	-	31	-	11	-	22
ER	8	54	-	20	-	11	-	-
Golgi apparatus	18	36	6	21	6	12	-	-
Endosomes-Lysosomes	30	30	15	11	-	7	7	-
Plasma membrane	34	23	17	11	8	-	-	-

A major difference between prokaryotic and eukaryotic cells is the presence of sterols in the membrane. The principal sterol present in mammals is cholesterol, which plays a fundamental role in modifying the main properties of the membrane such as fluidity, permeability, hydrophobicity and also promoting the formation of liquid-ordered (Lo) domains, mentioned before as lipid rafts. Cholesterol influences the fluidity of the membrane, and it does so in a bidirectional way. On the one hand, at high temperatures, cholesterol's flat, rigid structure limits phospholipid movement decreasing fluidity. This reduced fluidity can drive the formation of Lo domains enriched in cholesterol and saturated lipids (specially sphingolipids). On the other hand, cholesterol's kinked tail also disrupts lipid packaging at low temperatures preventing abrupt changes in membrane fluidity over a range of temperatures. Moreover, the amount of cholesterol in the membrane regulates the permeability by modifying hydrophobicity. High amount of cholesterol in the membrane greatly increases the activation energy required for polar and small ionic molecules to pass through the membrane core (Subczynski et al., 2017).

When speaking about biological membranes it is crucial to know the importance of the thermodynamics in protein insertion, folding and interaction. As it was mentioned before, biological membranes are a unique environment with very constrained physical chemical properties. For this reason, the amino acid composition of the membrane spanning polypeptide regions, the so called transmembrane (TM) segments, is critical for its proper insertion in the bilayer. However, before discussing the thermodynamics we need to introduce the concept of membrane proteins.

Membrane proteins are a group of macromolecules, that remains relatively unknown nowadays. It is estimated that in the human genome (but also in most genomes) ~30% of the genes are encoding integral membrane proteins (Almén et al., 2009; Krogh et al., 2001; Overington et al., 2006; Uhlén et al., 2015). Also, around 70% of all modern drug targets are membrane proteins (Kanonenberg et al., 2019). Despite that, among the approximately 8,000 membrane proteins identified in human cells by 2019, only around 50 have a determined high-resolution structure (Martin and

Sawyer, 2019). Analysing the RSCB Protein Data Bank (PDB) data (https://www.rcsb.org), it is possible to observe that membrane protein structures, even though their number has increased significantly (almost exponentially) in recent years, only represent around the 2% of all the PDB structures by 2022 (Figure 2), far away from their biological abundance and relevance. The explanation behind this phenomenon is the greater technical challenges associated with membrane protein studies. Structural analysis of proteins inserted in membranes used to be extremely challenging for X-ray crystallography, the principal (mainstream) approach for protein structural determination. Membrane proteins must be purified from disrupted membranes and replaced with carefully selected detergents for crystallization. In recent years, cryogenic electron microscopy (cryo-EM) single particle analysis (SPA) has become a convenient method for high resolution structural elucidation of membrane proteins, thus helping to solve gap of knowledge surrounding these proteins (Yao et al., 2020). One of the biggest problems for all the previous membrane protein isolation approaches was that they disrupt the membranes structures, hence abolishing any existing electrochemical gradient and membrane curvature. Structural analysis using electron cryo-electron tomography (cryo-ET) may be the ultimate solution to preserve these important properties to better connect protein structure knowledge and function (Lučić et al., 2008; Yao et al., 2020).
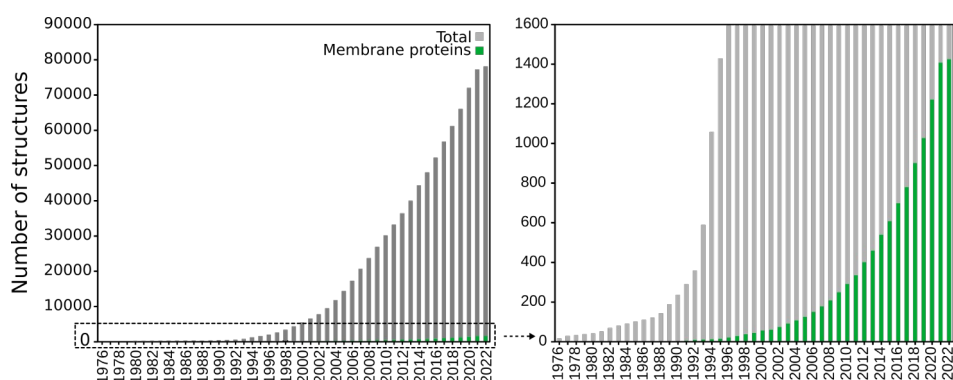
Figure 2. Comparison of membrane protein structures released in the PDB *versus* total (non-redundant) proteins. (**A**) Membrane protein unique structures (data from: blanco.biomol.uci.edu/mpstruc/) compared to total number of protein entries available on the PDB (non-redundant; data from: www.rcsb.org) among the

years 1976-2022. (**B**) Zoom for membrane protein unique structures from 1976 to 2022 (February).

Inside the big family of the membrane proteins, there are several possible classifications based on different properties. According to their membrane association, there are two main groups: the integral membrane proteins (IMPs) and the peripheral membrane proteins (PMPs). The IMPs are embedded in the lipid bilayer at different levels while the PMPs are associated reversibly but not integrated. Among the IMPs it is possible to distinguish between polytopic (or multi-spanning), bitopic (or single-spanning) and monotopic membrane proteins (non-spanning the lipid bilayer) (Figure 3) (Allen et al., 2019). As mentioned above, integral membrane proteins account for 20-30% of all genes in prokaryotic and eukaryotic organisms (Krogh et al., 2001). Most integral membrane proteins are formed by tight TM α-helix bundles (Hessa et al., 2007; Oberai et al., 2006). In other words, most membrane proteins are α-helical polytopic proteins, followed by bitopic and monotopic in that order. In the present thesis, we focused on integral membrane proteins, particularly on single- and multi-spanning membrane proteins formed by TM α-helical domains.
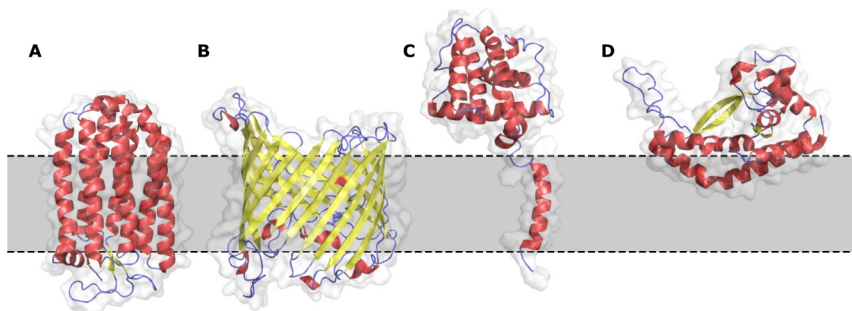
Figure 3. Structural examples of membrane protein classification. (**A**) A polytopic α-helical membrane protein (halorhodopsin; PDB code: 3QBG) (**B**) A polytopic β-sheet membrane protein (porin; PDB code: 1A0S) (**C**) Representation of a bitopic protein (BclxL) combining separate soluble and transmembrane structures (PDB codes: 3WIZ and 6F46, respectively) (**D**) Example of a monotopic membrane protein (pyridoxal 5'-phosphate (PLP)-dependent aminotransferase, RbmB; PDB code: 5W71; location calculated with PPM 3.0 server: https://opm.phar.umich.edu/ppm_server3).

TM α-helices are, on average, 24 amino acid residues long. Since the translation per amino acid in a canonical helix is 1.5 Å, a stretch of approximately 20 consecutive hydrophobic amino acids can span the 30 Å of the hydrophobic (apolar) core of a model biological membrane (Figure 4). Indeed, although the average is 24 amino acids, the most prevalent length of TM helices in membrane protein structures is 21 residues (Baeza-Delgado et al., 2013). As we described previously, the lipid bilayers into which this TM helices are inserted are at the same time highly hydrophobic, very anisotropic and with really constrained physical chemical properties (White and Wimley, 1999). This anisotropy is reflected in the distribution of different amino acids in the membrane-embedded parts of IMPs. Hydrophobicity is understood as a measure of the relative trend of a molecule to prefer a non-aqueous environment over an aqueous one. Therefore, in TM helices non-polar residues are prevalent, as it can be expected, in order to reduce the free energy (ΔG) of transferring amino side chains to the membrane hydrophobic core. One of the goals over the years has been unveiling which amino acids were more prone to membrane insertion than others. One of the most significant steps in this direction was the determination of the octanol-to-water solvatation free energies of acetyl amino acid amides published by White and co-workers in the mid-nineties (Wimley et al., 1996). Over the years, different amino acid hydrophobicity scales were developed in distinct solvents, serving as a starting point for TM domain polypeptide sequence prediction (MacCallum and Tieleman, 2011). One of the most relevant works in this field was the biological hydrophobicity scale proposed by von Heijne and collaborators (Hessa et al., 2005a). In this work, the authors depicted the contribution to the apparent free energy of membrane insertion (ΔG_{app}) for each of the 20 natural amino acids depending on the position of the residue in the TM sequence (Hessa et al., 2005a, 2007). Thanks to this scale, it is possible to predict the global ΔG_{app} for insertion of a helical sequence into the ER. The distributions of predicted ΔG_{app} values in native protein helices (Figure 5) enabled us to accurately distinguish helices from membrane and soluble proteins with better discrimination between data sets, compared to previous biophysical hydrophobicity scales (Hessa et al., 2007).
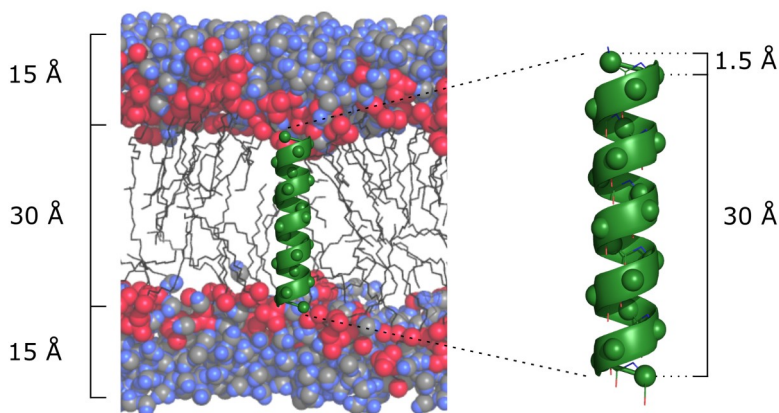
Figure 4. Schemiatic view of a TM α-helix in a biological membrane. Lateral view of a canonical α-helical TM domain (20 amino acids long, green) inserted in the hydrophobic core of model (DOPC) membrane. Distances of translation per residue and the full helix in Å are indicated. Notice that 20 amino acids in a canonical α-helical conformation have the precise length to span the membrane hydrophobic core (20 x 1.5 = 30 Å).
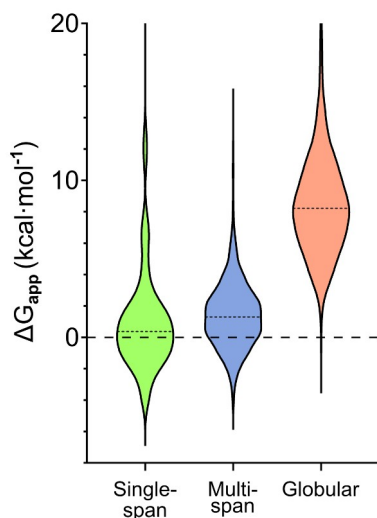
Figure 5. Distribution of predicted ΔG_{app} values for helices found in natural proteins with solved structure from PDB and PDBTM databases (non-redundant). Relative frequencies for helices from globular (red), multi-spanning (blue) and single-spanning (green) proteins according to their ΔG_{app} are shown. Helices from multi-spanning and single-spanning proteins show significantly lower ΔG_{app} values compared to globular ones.

More recent studies have continued this line of work by studying the coupling between insertion and interaction processes in the membrane milieu (Elazar et al., 2016). This topic will be further discussed in the following sections.

1.2 Targeting and insertion of α-helical membrane proteins in lipid bilayers

The protein folding has been recognized for a long time as a major conceptual problem to understand the information encoded in genes. Understanding protein folding in aqueous environments has been a primary goal for decades, while progress on membrane protein folding research lagged behind. Nonetheless, it was proposed that folding of membrane proteins may be more easily conceptualized than that of soluble proteins, because it takes place in a very constrained bidimensional space. One of the most influencing proposals on helical membrane protein folding was the so called two-stage model proposed by Popot and Engelman that divide this process into two energetically distinct stages: helix insertion and subsequent interactions among preinserted helices (Figure 6A) (Popot and Engelman, 1990). This simplified approach remains useful to explain and understand helical membrane protein folding and worked as a basis for more elaborated models (Popot and Engelman, 2000).

In the cellular context, membrane protein folding understanding requires to consider protein biogenesis, which can be divided into four processes: targeting, insertion, folding and assembly (Figure 6B) (Hegde and Keenan, 2022).
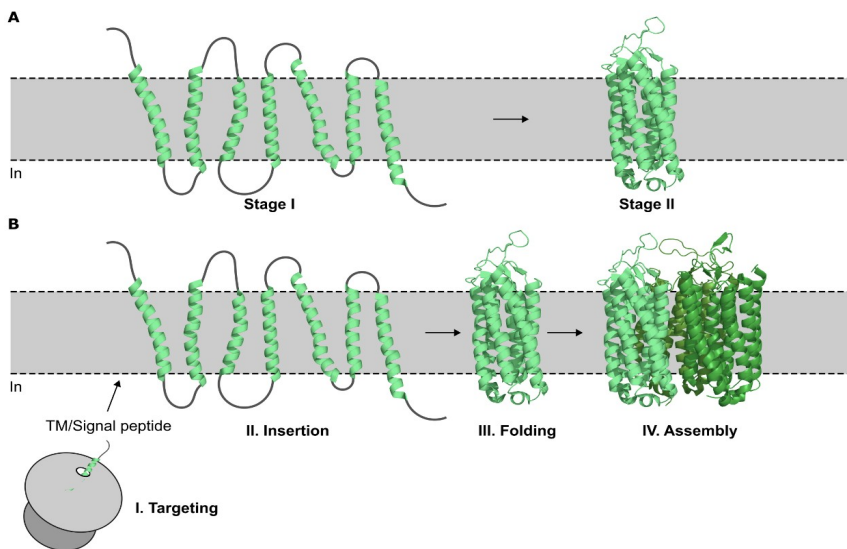
Figure 6. Schematic representation of membrane protein folding models. (**A**) Representation of the two-stages model for membrane protein folding proposed by Popot and Engelman. (**B**) The four major steps involved in membrane protein biogenesis.

Starting off with **targeting**, like all proteins, membrane proteins are synthesized by ribosomes, but in this case by membrane-bound ribosomes, that need to be guided to the membrane for their association. The targeting of a nascent chain exiting the ribosome is crucial to enable the delivery of these proteins to the membrane where they will be inserted. Membrane protein targeting to the eukaryotic ER or the bacterial plasma membrane is generally mediated by the most amino-terminal (N-terminal) hydrophobic domain within the newly synthesized protein. The targeting sequence can be either a cleavable hydrophobic sequence (most common) or the first TM domain of the protein (less frequent). Cleavable sequences, also known as *signal peptides*, can be cleaved off by an enzyme called "signal peptidase" after they have served their targeting function (Hegde and Keenan, 2022). The hydrophobic domain of signal peptides is around 7 to 9 amino acids (Almagro Armenteros et al., 2019) while TM domains are between 15 and 25 amino acids long (Baeza-Delgado et al., 2013; Krogh et al., 2001). This diversity in location and biophysical properties of the targeting sequence

underlines the requirement of different mechanisms or pathways to recognize and drive insertion of membrane proteins (Hegde and Keenan, 2022).

In eukaryotic cells, there are three main established targeting pathways depending on the position and hydrophobicity of the targeting sequence within the membrane protein sequence (Figure 7). These three routes are: the **s**ignal **r**ecognition **p**article (SRP) pathway, the **g**uided **e**ntry of **t**ail-anchored (GET) pathway and the **E**R **m**embrane protein **c**omplex (EMC) pathway. These three pathways can be differentiated in co-translational (SRP pathway) and post-translational (GET and EMC tail-anchored pathways) targeting (Hegde and Keenan, 2022; Whitley et al., 2021). In bacteria, most membrane proteins are targeted co-translationally by the SRP pathway or post-translationally using chaperones (Peschke et al., 2018). Because the eukaryotic ER is evolutionarily derived from the prokaryotic plasma membrane, the respective insertion machineries are related and share key mechanistic principles (Baum and Baum, 2014).

For co-translational targeting by the SRP, the most common pathway, a target sequence located at least ~65 amino acids from the carboxy terminus (C-terminus) of the nascent polypeptide is required (Figure 7A) (Zhang and Shan, 2014). This is because the SRP recognition occurs at the mouth of the ribosome exit tunnel, positioned around 35 amino acids from the peptidyltransferase centre inside the ribosome. Initially, the **n**ascent polypeptide-**a**ssociated **c**omplex (NAC) prevents SRP from binding to the ribosome. When a signal peptide or a TM domain emerges from the ribosome the NAC is exchanged for the SRP (Jomaa et al., 2022). After recognition, SRP-mediated targeting to the membrane takes between 5 to 7 seconds (Goder et al., 2000), during which another 30 amino acids can be synthesized (Hegde and Keenan, 2022). Co-translational targeting occurs only if translation does not terminate during this period, explaining why SRP requires targeting signals to be relatively away from the C-terminus. After the recognition, the SRP (together with the ribosome-nascent chain complex) will be guided to the SRP receptor in the membrane for the co-translational insertion via the Sec61 complex (Hegde and Keenan, 2022).

Membrane proteins which have only one TM segment closer than 65 amino acids from the C terminus are known as **t**ail-**a**nchored (TA) membrane proteins (Guna and Hegde, 2018). These proteins are mostly inserted post-translationally using chaperones (Figure 7B). The TM domains (TMDs) of this kind of membrane proteins differ widely in hydrophobicity, a key feature that will determine its targeting pathway (Guna et al., 2018). TA proteins are initially captured near the ribosome surface by the chaperone SGTA (Small Glutamine rich Tetratricopeptide repeat co-chaperone Alpha), whose recruitment to this site is facilitated by the GET4-GET5 complex. Those of higher hydrophobicity are then transferred to GET3 and finally to the GET1-GET2 complex in the membrane. Instead, those with lower hydrophobicity are not transferred to GET3, remaining soluble in the cytosol, and interacting with a wide variety of cytosolic chaperones (including SGTA, HSP70, ubiquitin family proteins and calmodulin) to be guided at the end to the cytosolic domain of the EMC in the membrane for their subsequent membrane insertion (Figure 7B) (Hegde and Keenan, 2022). The moderate hydrophobicity of the TMD is also a defining feature of mitochondrial membrane protein targeting. In this case, a network of cytosolic chaperones guides these proteins to the mitochondrial surface. This process appears closely linked to the protein targeting to the ER and causes mislocalization of mitochondria proteins to the ER and vice versa (Gupta and Becker, 2021).

This leads us to the second step of the process: **insertion**. Most membrane proteins in eukaryotic cells are recognized co-translationally at the ribosome by the SRP and targeted via the SRP receptor to the Sec61 translocon for insertion (Martínez-Gil et al., 2011; Voorhees and Hegde, 2015; Whitley et al., 2021). The Sec61 complex (analogue termed 'SecY complex' in bacteria and archaea) is a universally conserved heterotrimeric membrane protein that forms a protein-conducting channel used for both secretion and membrane insertion. The complex can open axially across the membrane for polypeptide translocation and laterally into the membrane for TMD insertion (Van den Berg et al., 2004). As mentioned above, there are

two different ways for the translocon to insert proteins in the membrane depending on their targeting: with or without signal-peptide.
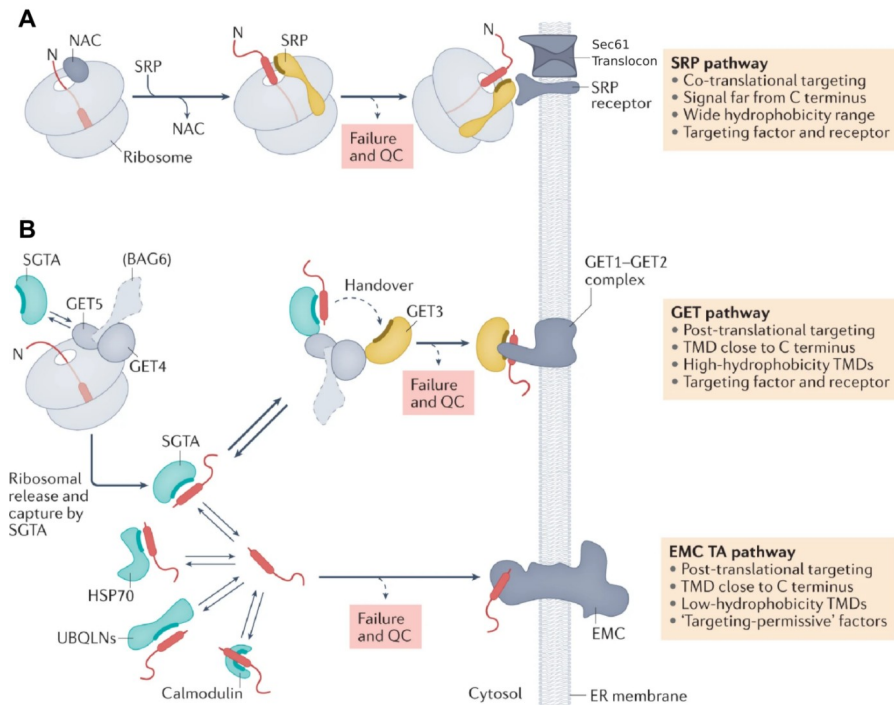
Figure 7. Membrane protein targeting to the endoplasmic reticulum. (**A**) The co-translational SRP pathway for delivering translating ribosomes to the ER membrane. Initially, the nascent polypeptide-associated complex (NAC) prevents SRP from binding to the ribosome. NAC is exchanged for SRP when a hydrophobic targeting sequence, either a signal peptide or a TMD, emerges from the ribosome. SRP then engages the SRP receptor at the ER (translocon-associated) to mediate targeting. (**B**) Post-translational targeting of TA proteins to the ER membrane. TA proteins are initially captured near the ribosome surface by the chaperone SGTA, whose recruitment to this site is facilitated by the GET4–GET5 complex. If the TMD of the TA protein is of high hydrophobicity, it is transferred to GET3 and then targeted to a receptor comprising the GET1–GET2 complex. If the TMD is of low hydrophobicity, it is instead kept soluble in the cytosol by cycles of binding and release from any of several chaperones. Subsequently, the TMD then engages the cytosolic domain of EMC. Adapted from Hegde and Keenan, 2022.

Membrane proteins engaged by GET3 are delivered to an ER-localized receptor composed by membrane bound GET1-GET2 complex, which is necessary and sufficient for its insertion. This insertion process

involves three main steps: initial engagement of TA-GET3 targeting complex by the GET1-GET2 receptor, receptor-mediated release of TA protein from GET3, and TM domain insertion into the bilayer (Schuldiner et al., 2008). The EMC, as said before, is a more recently described possible pathway for TA protein insertion in the ER (Guna et al., 2018). Sometimes, natural TM domains can have lower hydrophobicity than what is efficiently accommodated by GET3. These discrepancies were explained, at least partially, with the discovery of the widely conserved protein complex termed EMC that can insert TA-proteins with low and moderate hydrophobic TM segments (Hegde and Keenan, 2022; Jonikas et al., 2009). **Folding** and **assembly** will be discussed in section 1.4.
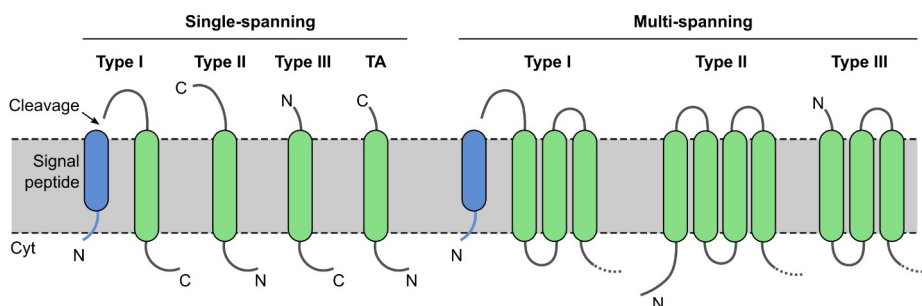
Figure 8. Topological classification of integral membrane proteins. Major classes of integral membrane proteins are indicated. C, carboxy terminus; N, amino terminus.

From the biophysical point of view, the hydrophobicity and thickness of the core of the membrane bilayer require to minimize polarity of TM segments for a proper insertion. One of the principal sources of polarity in polypeptides is the peptide bond itself. To minimize the cost of harbouring a polar polypeptide backbone in TM segments, it is expected for them to adopt a regular secondary structure (mainly α-helical) maximizing a regular pattern of hydrogen bonds between backbone carbonyl (C=O) and amide (NH) groups to significantly reduce the free energy penalty but also to fit in length. In helical TM domains this secondary structure is already achieved in the ribosome exit tunnel before entering the translocon (Bañó-Polo et al., 2018). Once in the translocon channel, a nascent polypeptide chain has to either insert into the hydrophobic membrane core or remain in the aqueous environment for further translocation. This dichotomy appears to be largely

governed by the thermodynamic preference of the lateral chains for residing in each environment. The thermodynamics are dominated by the hydrophobic effect, which depends on how well the physical-chemical properties of the amino acid stretch can match the properties of the lipid membrane (Corin and Bowie, 2022; Cymer et al., 2015). To understand these amino acid properties, numerous hydrophobicity scales have been developed to quantify the probability of insertion for the 20 natural amino acids in the membrane. As commented in the previous section, initial scales were modelled based on the transfer free energy between solvent and water partition (Wimley et al., 1996), while more membrane-specific and biologically relevant scales were subsequently developed (Hessa et al., 2007, 2005a). Hessa's scale was based in an *in vitro* assay using microsomal membranes, then quantifying insertion propensities in a biological context. These authors measured the insertion probability using a large series of 19 amino acid long test segments. This biological hydrophobicity scale correlates with previous biophysical hydrophobicity scales but reflects the apparent transfer free energy from the translocon to the bilayer. This scale was later refined to describe the contribution of each side chain at each position along a TM helix (Hessa et al., 2007).

Although hydrophobicity scales can predict how prone to insertion a helix is, they lack some information. For example, as it can be seen in Figure 5, sometimes multi-spanning membrane proteins helices are only marginally hydrophobic and do not insert efficiently on their own. Yet, they are still able to properly insert and fold in the membrane suggesting that other factors can facilitate the insertion of polar sequences in the membrane (Corin and Bowie, 2022). The position of neighbouring helices and the interaction with other proteins can influence TM helix insertion. Polar residues are way more prevalent than expected in the membrane. Inter-helix interactions between polar residues can reduce the energetic barrier for polar residues insertion, not expected to properly accommodate in the hydrophobic core on their own (Bañó-Polo et al., 2013; Illergård et al., 2011). Also, intra-helical salt bridges between polar residues have been previously suggested as a possible mechanism to ameliorate the penalty for inserting functional or key charged residues present in TM segments (Bañó-Polo et al., 2012).

Similarly, TM helix length can also affect insertion. Compared to water-soluble proteins, TM helices must satisfy the demanding restrictions imposed by the complex membrane environment (Baeza-Delgado et al., 2013). Helix lengths in TM proteins typically ranges from 14 to 36 residues, although several experimental and computational studies suggest that a minimum of around 10 leucine residues are required for efficient insertion (Baeza-Delgado et al., 2016; Jaud et al., 2009). The higher energetic cost of inserting shorter helices is likely due to hydrophobic mismatch. While the hydrocarbon core of a bilayer is around 30 Å, a 10-residue canonical helix only spans ~15 Å requiring the protein or the bilayer (or both) to deform with a subsequent energetic cost (Corin and Bowie, 2022). Nevertheless, severe membrane-helix rearrangements have been observed in cellular membranes and in molecular dynamics (MD) simulations to accommodate fairly large mismatches (Grau et al., 2017).

In summary, as it is expected, most apolar amino acids (Ala, Leu, Met, Ileu, Phe, Val) are more probable to locate in the apolar hydrocarbon core region, while polar and charged amino acids (Asp, Asn, Glu, Gln, Lys, Arg) are preferentially found near the surface surrounded by polar lipid headgroups and bulk water. Of these, Lys and Arg (positively charged) typically show a strong preference for the flanking cytoplasmic side of the bilayer, reflecting the so-called "positive inside rule" which helps to determine membrane protein topology (Nilsson et al., 2005; von Heijne, 1989). This is the main topological determinant that we find in TM domains, but not the only.

1.3 Topological determinants

In membrane proteins, the term *topology* refers to the two-dimensional information that indicates how many TM domains are in the membrane and the orientation of soluble domains relative to the plane of the membrane. Although membrane protein topology is mostly determined by its biogenesis and insertion mechanism, there are some topological determinants that play an important role in this dynamic process.

It has been observed that positively charged residues flanking TM domains tend to have an asymmetrical distribution in membrane proteins. Positive charges are predominantly found in the cytoplasmic side of the membrane, causing as, a result, a phenomenon known as ***positive-inside rule*** (von Heijne, 2006, 1989, 1986). The positive-inside rule, firstly proposed by von Heijne, has been confirmed with statistical observation for most membrane proteins and biological membrane types (Baeza-Delgado et al., 2013; Baker et al., 2017). There are many possible explanations to this rule. Positively charged residues in a nascent chain seem to be less prone to translocate across the membrane and are left in the cytosolic side of the membrane where bound ribosomes are placed and protein translation occurs. Since membrane proteins are mainly co-translationally translocated and inserted in the membrane by the translocon, positively charged residues exposed to the inner side of the membrane may interact with the negatively charged lipid head groups hindering translocation. Furthermore, the translocation machinery may not be able to accommodate positive charges easily. Indeed, in the yeast Sec61p translocon subunit negatively charged residues (Glu382) lie near the cytoplasmic end of the channel while positively charged residues (Arg67 and Arg74) lie closer to the other end of the channel (Bogdanov et al., 2014). These residues are also conserved in mammals (Glu381, Arg66 and Arg73 in humans). Lastly, the negative membrane potential in the cytoplasmic side of the membrane may facilitate the accommodation of positively charged residues, while the positive potential in the other side of the membrane would repulse positively charged residues across the membrane (Baker et al., 2017; Lee and Kim, 2014). In this last situation, the asymmetric distribution of lipids in the leaflets of the plasma membrane can play an important role. Specifically, negatively charged phosphatidylserine (PS) was found to distribute preferentially in the cytosolic side of the membrane, and it was found to electrostatically interact with moderately positive-charged proteins enough to redirect the protein localization (Yeung et al., 2008). Numerous studies show that addition or deletion of positively charged flanking residues strongly affect the orientation of the TM domain according to the positive inside rule (Lee and Kim, 2014), being possible to invert topology by only modifying the charged

flanking region of a TM domain (Goder and Spiess, 2001; von Heijne, 1989).

In bacterial membrane proteins the positive-inside rule is better preserved than in eukaryotic membrane proteins. In fact, it is rare to find positively charged residues in soluble loops facing the periplasm, while in eukaryotic membrane proteins the soluble domain oriented to the cytoplasmic side contains a net positive charge but it is common to find positively charged residues in the outer side of the membrane. This difference have been attributed to the differences in the membrane potential of the different membranes (Lee and Kim, 2014). Engagingly, the positive-inside rule observed in most membrane proteins does not exhibit a noticeable effect in nuclear-encoded mitochondrial inner membrane proteins (Gavel and von Heijne, 1992). These proteins are translocated from the cytosol/intermembrane space by interacting with lipid head groups. Since the inner membrane potential is negative in the matrix side, the electrophoretic force may facilitate translocation of positively charged residues across the inner membrane. This results in positively charged residues equally distributed in the intermembrane space and the matrix of the mitochondria (Lee and Kim, 2014).

Because of the reasons exposed previously, it may be expected for negative charges to also be asymmetrically distributed in the TM domains nearby regions. It might be inferred that, if positive residues force the flanking loop or tail of a TMD to stay inside, negative residues would be drawn outside. Although acidic residues are rare within and nearby TM segments, there has been proposed a ***negative-inside depletion/negative-outside enrichment rule*** (also *negative-not-inside/negative-outside* rule) that complements the positively charged bias (Baker et al., 2017). Sequence statistical studies have shown that here is a less preserved trend for negatively charged residues to prefer the outside flank of a TM helix rather than the inside. This trend is stronger in single-spanning protein datasets than in multi-spanning. It was suggested that net zero charged neutral phospholipid (e.g.: PE) dampen the translocation potential of negatively charged residues in favour of the cytoplasmic retention potential of

positively charged residues. This explains why positively charged residues are more potent topological signals than negatively charged residues (Bogdanov et al., 2014). While the "negative-not-inside/negative-outside" bias is observed for distantly related eukaryotic species, and it is also present in Gram-negative bacteria such as *E. coli*, it seems that is not a universal trend since this pattern was not observed in Gram-positive bacteria (Baker et al., 2017).

There is also a statistical correlation between the net electrical charge difference between the extramembrane domains flanking the TM segments. This *charge difference rule* differs from the positive-inside rule by giving positive and negative charges the same topological strength (Bogdanov et al., 2014). Even though it has been demonstrated that negative charges have less influence over the topology determination (Baker et al., 2017; Nilsson and von Heijne, 1990), the complementation between the positive-inside rule and the negative-inside depletion/negative-outside enrichment rule can be understood as a charge difference balance rule guiding membrane protein topology.

1.4 Folding, assembly and interaction motifs

Multi-spanning membrane protein **folding** is often driven by the presence of non-hydrophobic amino acids (Figure 9A and B). Thus, successful folding of multi-spanning membrane proteins requires the biogenesis machinery to not only insert partially hydrophilic TMDs but also temporarily stabilize and shield them in the membrane until their proper assembly with other TMDs avoiding the ER-associated quality control degradation pathways (Hegde and Keenan, 2022). Calculations of insertion apparent ΔG suggest that single-spanning TM segments are typically more hydrophobic than TM segments of multi-spanning membrane proteins (Hessa et al., 2007). This tendency is also observed in the distribution of predicted ΔG_{app} values for helices found in natural membrane proteins with solved structure from PDBTM databases (Figure 5). Likely, folding of multi-spanning IMPs can help to better bury polar residues in the core of the tertiary structure. Actually, polar residues are more prevalent and conserved than expected in

the membrane core of α-helical membrane proteins. These residues constitute as much as 9% of all residues in TMD from protein structures, and same fraction is predicted to exist within all membrane proteins of the human genome (Illergård et al., 2011). Charged residues can promote folding by association because of the hydrophobic effect. Also, salt bridges formation between membrane-spanning charged residues is required to promote TMD-TMD interactions during folding in some membrane proteins while lowering the unfavorable energetics of inserting charged residues into the membrane (Bañó-Polo et al., 2013, p.). The mechanisms of multi-spanning membrane protein biogenesis and the stabilization of hydrophilic residues in the membrane are just beginning to be defined (Hegde and Keenan, 2022).
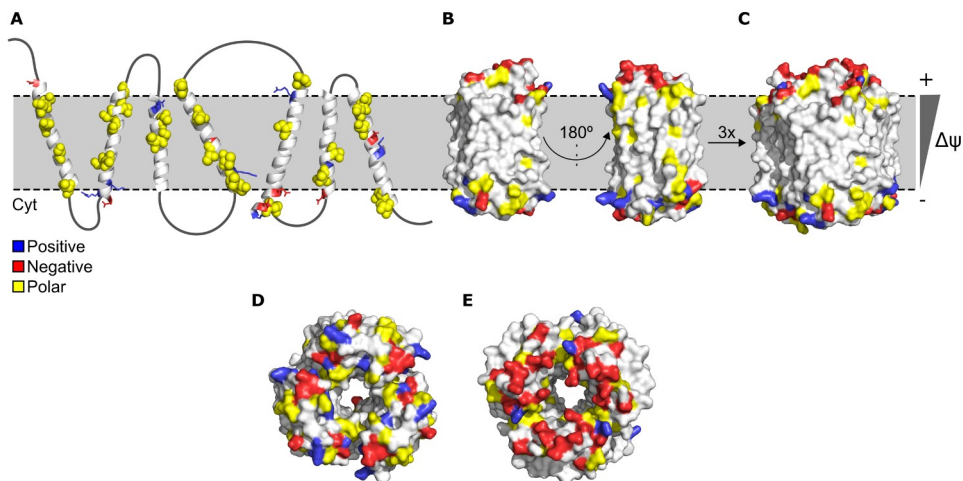
Figure 9. Insertion, folding and association of a multi-spanning membrane protein. The linear (**A**), folded (**B**) and multimeric (**C**) states of *Natronomonas pharaonis* halorhodopsin (PDB ID: 3QBG) illustrate how most of the hydrophilic side chains (positively, negatively and polar shown in blue, red and yellow respectively) in the transmembrane domain (TMD) become buried during folding. Also, the charge distribution can be used as an example of the positive-inside rule and the negative-inside depletion/negative-outside enrichment rule. (**D**) Downside view of the quaternary structure of halorhodopsin from the cytosol side. (**E**) Top view of the quaternary structure of halorhodopsin from the extracellular side.

One of the leading hypotheses to explain helical membrane protein biogenesis was the existence of chaperoning proteins associated to the

translocon. Pioneering experiments using crosslinking found a glycoprotein known as TRAM (**Tr**anslocation **A**ssociated **M**embrane protein) to be associated to TM sequences during cotranslational integration of a nascent protein into the ER membrane (Do et al., 1996; Görlich et al., 1992). TRAM showed to wield some restrictions on TMD movement, not allowing lateral diffusion of client nascent chains until ending of translation (Do et al., 1996). This movement restrictions may allow the client protein to properly fold before being exposed to the membrane environment. Furthermore, this protein is present in sub-stochiometric amount and may only associate transiently under certain circumstances when particular substrates are present in the translocon (Whitley et al., 2021). Therefore, TRAM has been proposed to perform a TM chaperone-like role during the integration of nonoptimal TM segments into the bilayer (Tamborero et al., 2011). More recent experiments using chemical crosslinking to find proteins adjacent to various insertion intermediates of the seven-TMD protein rhodopsin found, in addition to the ribosome associated Sec61 complex, an unidentified ~10 kDa protein provisionally termed PAT10 (**P**rotein **A**ssociated with the **T**ranslocon of **10** kDa) (Hegde and Keenan, 2022; Ismail et al., 2008; Meacock et al., 2002). This protein was recently identified to be part of an intra-membrane chaperone complex that facilitates membrane protein biogenesis (Chitwood and Hegde, 2020). This complex, known as the PAT complex, is formed by an abundant obligate heterodimer of CCDC47 and Asterix (previously identified as PAT10). The PAT complex binds to nascent TMDs containing lipid exposed hydrophilic residues and disengages during substrate folding. Cells lacking either subunit of the PAT complex show reduced biogenesis of numerous multi-spanning membrane proteins (Chitwood and Hegde, 2020).

Also, a large (~360 kDa) ribosome associated complex comprising the core Sec61 channel and five accessory factors (TMCO1, CCDC47 and the Nicalin-TMEM147-NOMO complex) has been recently described (McGilvray et al., 2020). The absence of Asterix, the obligate partner of CCDC47, in this proteomic experiment is probably explained by technical limitations associated with the paucity of Asterix-derived peptides generated by trypsin digestion (Chitwood and Hegde, 2020; Hegde and Keenan, 2022).

High-throughput sequencing of the mRNAs recovered with this complex shows selective engagement with hundreds of different multi-pass membrane proteins. This information suggests the presence of a specialized human translocon for multi-spanning membrane proteins and provides a molecular framework for understanding its role in membrane protein biogenesis (McGilvray et al., 2020). The mammalian multi-spanning translocon may be analogous to the prokaryotic complex known as "holotranslocon". The lipid-filled cavity of the mammalian multi-spanning translocon is large enough to accommodate multiple TM segments with their exposed hydrophilic parts being temporally chaperoned, offering a protected site for membrane protein folding (Hegde and Keenan, 2022).

After proper folding of membrane proteins (or straight after insertion in case of single-spanning), many newly folded proteins interact with previously embedded membrane proteins. Around a half of the membrane proteins are part of multiprotein complexes, many with other membrane proteins. The mechanism of **assembly** of two or more membrane proteins within the lipid bilayer is poorly understood. The simplest model suggests that the individual subunits diffuse within the membrane until they encounter their partner or partners. This mechanism, although plausible, is unlikely to be the primary strategy used in a crowded cellular environment (Hegde and Keenan, 2022). Because of the presence of quality control systems that recognize aggregation of unassembled orphan units, the presence of assembly chaperones to guide the assembly has been suggested although they are not well defined.

Single-span TM helices have structural and functional roles well beyond serving as mere anchors for soluble domains in the membrane. They frequently direct the assembly of protein complexes and mediate signal transduction. There are not many identified motifs for helix-helix interaction. The first described motif was the GxxxG (being G the Gly and x any other amino acid) identified in human glycophorin A (GpA) TM domain. The GxxxG motif appears to frequently mediate and stabilize TM helix-helix interactions in membranes (Brosig and Langosch, 1998; Finger et al., 2009; Lemmon et al., 1992b, 1992a). This motif can be extended to small-xxx-

small (considering small Gly, Ala, Ser or Thr) which can also mediate and stabilize defined TM helix-helix interaction (DeGrado et al., 2003; Finger et al., 2009; Moore et al., 2008). Small residues are believed to allow two helices to come into close contact so that other forces, such as van der Waals interactions and hydrogen bonding, can stabilize a given TM helix bundle (Finger et al., 2009; Martinez-Gil and Mingarro, 2015). In some cases, even a single glycine residue appears to be sufficient as a framework for TM helix-helix interactions (Finger et al., 2009). Notice that in the GxxxG the three amino acids between the two Gly allow a complete helix turn, facing both Gly in the same bidimensional helix lateral surface. This pattern can be repeated, always with the constant of having small residues in the same face of the helix resulting in a Gly zipper (or a *small residue* zipper). Other motifs like the SxxSSxxT and SxxxSSxxT, obtained by Engelman and collaborators from a randomized library, also respond to the same distribution (Dawson et al., 2002; Moore et al., 2008). These interaction motifs can drive parallel and antiparallel interactions (with left-handed and right-handed interactions in both cases) between helices, giving as a result homo and hetero-oligomers. Most known interacting monotypic TM helices form parallel oligomers. Parallel right-handed pairs commonly have a ~40° crossing angle and a GxxxG-like motif, while parallel left-handed helices have a ~20° crossing angle and often a heptad repeat of small residues (Moore et al., 2008; Walters and DeGrado, 2006). Also, heptad repetitions of Leu residues can lead the self-assembly of TM segments in a structural equivalent to soluble leucine zipper interaction domains (Gurezka et al., 1999; Ruan et al., 2004).

Small residues are not the only prevalent side chains in TM interacting interfaces. Strongly polar side chains can also play an important role in the association of TM domains in a variety of packing motifs. For example, a Thr residue at the helix-helix interface of GpA appears to cooperate with the GxxxG motif to increase the affinity and specificity of the interaction (Lemmon et al., 1992b; Moore et al., 2008). In other cases, a single polar residue can play a dominant role in stabilizing helix-helix oligomerization. Model peptides containing a single Asn, Asp, Gln or Glu oligomerize in detergents, liposomes and bacterial membranes (Choma et al., 2000; Moore et al., 2008; Ruan et al., 2004; Zhou et al., 2000).

All the interaction motifs described above can drive stable (assembly of quaternary structures) but also transitory interactions. As previously described, TM domains do not only function as anchors for soluble domains in the membrane. Even though sometimes they may have a strictly structural function, IMP are the main vehicle for communication between extracellular and intracellular media, but also between inner and outer spaces for organelles in eukaryotic cells. Interactions within the membrane span regulate complex responses and signalling pathways; thus, interactions must be dynamic but tightly regulated. These interactions, therefore, can be used as a target for controlling different processes. Although modelling, design and engineering of membrane proteins has advanced significantly over the past twenty-five years (Mingarro et al., 1997), they still lie far behind those of soluble proteins (Weinstein et al., 2019). The design of intra-membrane interactions hardly has two successful precedents with the seminal computational design of membrane-integral inhibitors fifteen years ago (Yin et al., 2007) and recent *de novo*-designed transmembrane domains that tune engineered receptor functions (Elazar et al., 2022). In essence, membrane proteins are an exciting and promising research area that still needs to be further explored.

2. Objectives

The main objective of this thesis is to increase our understanding of membrane protein biogenesis from insertion and topology determination to the importance of TMD-TMD interactions for regulation of different cellular processes. The following specific objectives have been addressed along the thesis:

- To study the formation of intra-helical salt bridges and to estimate their contribution to TMD insertion into biological membranes.

- To assess the influence of extramembranous charged residues on membrane protein topology determination.

- To comprehend the importance of intramembrane interactions in the regulation of complex processes like viral controlled apoptosis.

- To evaluate the use of TMD-TMD interactions as therapeutic targets.

- To computationally/rationally design intramembrane interactions

3. Results and discussion

The main objective of this thesis was to study α-helical membrane protein segments, from basic principles to rational design. Our intention was to investigate a wide spectrum of functions developed by the TMDs far beyond from its structural role anchoring proteins to membranes. The final goal of this thesis is to delve deep in our understanding of membrane protein biogenesis from insertion and topology determination to better characterize these 'greasy' proteins. In particular, we wanted to make progress on depicting how α-helical TMD-TMD interactions work and their importance in the regulation of different cellular processes like apoptosis and the control of cellular death. A better understanding of these regulation processes can improve our current knowledge and lead to new therapeutic targets. In this direction, we ought to explore the inhibition of TMD-TMD interactions as a target to regulate cell death.

The results of this thesis were obtained using some of the more relevant techniques from the biochemistry and molecular biology toolbox as: plasmid cloning, site-directed mutagenesis, SDS-PAGE electrophoresis, western blotting, transcription/translation expression *in vitro*, apoptosis assays in cell lines (HeLa), and also protein engineering approaches based on bimolecular complementation protocols like BiFC (**Bi**molecular **F**luorescence **C**omplementation) and BLaTM (based on the use of a split **β-La**ctamase fused to a **TM**D).

These techniques led us to a better description of how polar residues can be inserted into the membrane. They also helped us to improve our understanding of the insertion process, the topology determination of viral proteins and to better comprehend the complexity of TMD-TMD interactions and their role to fine-tune the apoptotic networks, pointing out the importance of the TM region. Finally, we used these interactions to computationally design TM inhibitors as a new possible therapeutic agents to regulate cellular death by modulating TMD-TMD interactions. The research related to this work resulted in four first-author publications and one manuscript that is currently being revised, which constitute the four chapters

of this thesis, and two more papers included in the annexed section. The annexed papers (see *section 8*), whereof I am the first author, include a review depicting some of the most relevant methods used in this thesis to study TMD interactions. Additionally, a patent (P202230029) including part of this work has been filled by the University of Valencia. The following pages are a summary of the results and discussion of these papers organized in four chapters.

3.1. Chapter 1: Insertion

Publication:

Duart, G.*, Lamb, J.*, Ortiz-Mateu, J., Elofsson, A., Mingarro, I., 2022. *Intra-Helical Salt Bridge Contribution to Membrane Protein Insertion.* **Journal of Molecular Biology** 434, 167467. https://doi.org/10.1016/j.jmb.2022.167467
*Equal contribution

The objective of this work was to estimate the energetic contribution of intra-helical salt bridges to the insertion of TMDs into biological membranes. The presence of intra-helical salt bridges in TMDs, as well as their impact on insertion, has not been properly studied yet. α-helical TMDs are largely composed of apolar residues because of the hydrophobic nature of the membrane. Nevertheless, in some cases, membrane embedded proteins carry polar amino acids in their TMDs for proper folding or functional purposes (Baeza-Delgado et al., 2013). In fact, the presence of polar and charged amino acids in TMDs is more frequent than what would be expected according to the hydrophobic nature of the environment (Bañó-Polo et al., 2012), especially when these residues are present in pairs.

Salt bridges are electrostatic interactions between negatively and positively charged amino acids that play a prevalent role in protein stabilization (Marqusee and Baldwin, 1987). Analysing the TMDs from membrane protein structures we observed that charged pairs of amino acids are especially prevalent at positions *i, i+1*; *i, i+3* and *i, i+4*. Oppositely paired charges located at these positions could potentially form salt bridges

as they are all on the same face of the helix and are close enough, in terms of atomic distances.

To assess the contribution of putative salt bridges to the translocon-mediated membrane insertion, we used as a vehicle the leader peptidase (Lep) protein from *Escherichia coli*. The Lep protein consists of two TMDs (H1 and H2) connected by a cytosolic loop (P1) and a large C-terminal (P2) domain, which inserts into endoplasmic reticulum (ER)-derived rough microsomes (RM) with both termini located in the lumen of the microsomes (von Heijne, 1989). The designed TMDs were inserted into the luminal P2 domain and flanked by two *N*-linked glycosylation acceptor sites (G1 and G2). Glycosylation occurs exclusively in the lumen of the ER (or the microsomes) because of the location of the oligosaccharyltransferase (OST) active site (a translocon-associated enzyme responsible for the oligosaccharide transfer) (Braunger et al., 2018). Glycosylation of an acceptor site increases the apparent molecular mass of the protein (~2.5 kDa), which facilitates its identification by gel electrophoresis.

We first compared the effect of the oppositely charged Lys and Asp residues on the insertion of a 19-residue-long hydrophobic artificial scaffold (L4/A15, 4 leucines and 15 alanines), designed to stably insert into the RM membranes and "insulated" from the surrounding sequence by N- and C-terminal GGPG- and -GPGG (G, glycin; P, prolin) tetrapeptides. Single Lys and Asp residues were placed at positions 8 and 12 respectively, and pairs of Lys-Asp residues were designed to cover positions 7–12 (more than one helical turn). When pairs of charged residues were present, our results showed a tendency to insert more efficiently when oppositely paired charges were placed at positions that are permissive distances for salt bridge formation (*i, i+1*; *i, i+3*; *i, i+4*), an effect not observed in the current prediction algorithms. Similar results were obtained on a different Leu/Ala background with a slightly higher insertion efficiency (L5/A14, 5 leucines and 14 alanines).

After testing the effect of oppositely charged residues in the insertion of model sequences, we decided to look for salt bridges in natural

proteins. We analysed the TMDs from high-resolution membrane protein structures. Then, we generated a list of potential candidates for further *in vitro* and *whole cell* studies. We selected the helix G from *Natronomonas pharaonis* halorhodopsin (PDB code: 3QBG) and the helix A from the *Oryctolagus cuniculus* calcium ATPase (PDB code: 1SU4). In both cases we studied the insertion of these helices using *in vitro* assays (Lep in microsomes) and eukaryotic cells (Lep and C_LTM systems).

Halorhodopsin from *Natronomonas pharaonis* is a protein made up of seven TMDs (A to G) and a retinal chromophore bound via a protonated Schiff base to the ε-amino group of the Lys258, located in the middle of helix G (Kanada et al., 2011). *In silico* analysis of the apoprotein 3QBG structure showed a Lys-Arg pair, involving *i, i+4* Lys258 and Asp254 residues from helix G. The distance between charges in the crystal structure of the apoprotein was about 3.5 Å, a permissive distance for a salt bridge formation. Then, we designed three mutants that were supposed to perturb salt bridge interaction in different ways: the K258D mutant places two charged residues with the same polarity at positions *i, i+4*; the K258A mutant replaces one of the charged residues by a non-polar amino acid, and the K258Y/Y259K double mutant places the two different charges at a non-permissive position for salt bridge formation (*i, i+5*) while preserving amino acid composition. The results of the Lep-based glycosylation assay indicated that wild type and K258A mutant are inserted properly, but when the salt bridge is disrupted, insertion efficiency decreased substantially (~0.5 kcal/mol). These results were replicated in HEK-293T cells. Cells transfected with the chimera containing helix G native sequence rendered almost full insertion. In contrast, cells transfected with the construct harbouring *i, i+5* sequence rendered almost exclusively membrane translocation. These results emphasized the relevance of intra-helical salt bridges in translocon-mediated TM insertion, especially in the cellular environment.

Our second studied protein was the calcium ATPase from *Oryctolagus cuniculus*. This protein contains a bundle of 10 TM helices (A to J) (Toyoshima et al., 2000). *In silico* analysis of 1SU4 highlighted an Asp-

Arg pair, involving Asp59 and Arg63, in the centre of helix A. The distance between these charges in the crystal structure was about 3.0 Å, clearly within the permissive range for salt bridge formation. Again, the results of the Lep-based glycosylation assay demonstrated the efficient insertion of the native helix A sequence. Importantly, when the paired residues were placed at the non-permissive *i, i+5* positions the insertion efficiency was decreased, with a ΔG_{app} estimated in ~0.7 kcal/mol. Similar results were also observed in HEK-293T cell experiments.

In summary, charged residues found in α-helices can be important for function (Lin and Lin, 2018) but also provide a stabilizing effect (Armstrong and Baldwin, 1993). Pairs of oppositely charged residues can form salt bridges that could allow them to facilitate TMD insertion when facing the hydrophobic environment (Walther and Ulrich, 2014). Salt bridges might also be important for shielding the charges during translocon-mediated TMD insertion (Whitley et al., 2021). In TMDs, oppositely charged residue pairs are more prevalent at position*s i, i+1; i, i+3* and *i, i+4,* compatible with salt-bridge formation. By analysing two native helices containing intra-helical salt bridges we found that the free energy of insertion (ΔG_{app}) is significantly reduced when both oppositely charged residues are spaced at a permissive distance. These results indicate that intra-helix salt bridges could form during translocon-assisted insertion or even earlier, since TM helices can be compacted inside the ribosome exit tunnel (Bañó-Polo et al., 2018). The reduction of ΔG_{app} in these natural proteins is between 0.5-0.7 kcal/mol. As found in the case of the halorhodopsin helix G, this reduction might be higher in the cell context, since some auxiliary components of the membrane insertion machinery (Chitwood and Hegde, 2020; Shurtleff et al., 2018; Tamborero et al., 2011) might be not represented in the microsomal vesicle preparations. Current prediction algorithms for membrane protein insertion tend to overestimate the free energy penalty of charged residues in TMDs. Incorporating the effect of potential salt bridges in the reduction of ΔG_{app} during membrane integration could help to improve future prediction tools.

3.2. Chapter 2: Topology

Publications:

Duart, G.*, García-Murria, M.J.*, Grau, B.*, Acosta-Cáceres, J.M., Martínez-Gil, L., Mingarro, I., 2020. *SARS-CoV-2 envelope protein topology in eukaryotic membranes.* **Open Biology** 10, 200209. https://doi.org/10.1098/rsob.200209
*Equal contribution

Duart, G., García-Murria, M.J., Mingarro, I., 2021a. *The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness.* **Biochimica et Biophysica Acta (BBA) - Biomembranes** 1863, 183608. https://doi.org/10.1016/j.bbamem.2021.183608

As soon as the genome of the SARS-CoV-2, the virus responsible of the COVID-19 pandemic, was released, we started a project aimed to study the envelope (E) protein. The E protein is the smallest and has the lowest copy number among the membrane proteins found in the lipid envelope of mature virus particles (Bar-On et al., 2020). However, it is critical for pathogenesis of the SARS-CoV-2 and other human coronaviruses (Almazán et al., 2013; Ruch and Machamer, 2012; Xia et al., 2021; Zheng et al., 2021) and has been described as a viroporin. 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 (Wu et al., 2020). This sgRNA encodes a 75 residues long polypeptide with a predicted molecular weight of approximately 8 kDa. 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, with only the initial Met, Leu39, Cys40 and Pro54 being ubiquitously conserved. Regarding overall sequence similarity SARS-CoV-2 E protein has the highest similarity to SARS-CoV (94.74%) with only minor differences, followed by MERS-CoV (36.00%). Interestingly, sequence similarities are significantly lower for the other four human coronaviruses, which usually cause mild to moderate upper-respiratory tract illness typical for common cold.

To determine its membrane topology, we assayed E protein insertion in microsomal membranes using *in vitro* transcription/translation experiments in the presence of [^{35}S]-labelled amino acids but also in eukaryotic membranes using HEK-293T cells. Using a glycosylation based assay as a molecular reporter, we determined that the SARS-CoV-2 E protein integrates into the membrane co-translationally as a single-spanning membrane protein with an Nt_{lum}/Ct_{cyt} orientation in *in vitro* and *in vivo* systems. This topology is compatible with the ion channel capacity described previously (Verdiá-Báguena et al., 2012). Furthermore, this topology is reinforced by different topological determinants present in the SARS-CoV-2 E protein sequence. In all seven human coronaviruses there is a strongly conserved positively charged residue placed after the hydrophobic region. It is worth to mention that this residue is an arginine (Arg38) in MERS-CoV, SARS-CoV and SARS-CoV-2, while in the other human coronaviruses is a lysine. Also, the alignment of MERS-CoV, SARS-CoV and SARS-CoV-2 E proteins unveils a tendency to accumulate a net positive charge balance C-terminally to the TMD, which correlates with the positive-inside rule, suggesting an increasing robustness in the topology determination from MERS-CoV to SARS-CoV-2. We experimentally confirmed this increasing robustness by modifying the full protein charge balance for all three pathogenic E proteins. In all three cases, the conserved Arg38 residue plays a limited role in the topology determination. Our data also suggested that the Arg to Glu mutation present in both SARS-CoVs' N-terminus compared with MERS-CoV sequence, is most likely one of the mechanisms contributing to the proved topology robustness of the SARS-CoVs by converting the net charge of 0 at N-terminal region of MERS-CoV into a -2 in both SARS-CoVs, in good agreement with the so-called "negative outside enrichment" rule (Baker et al., 2017).

3.3. Chapter 3: Interaction and function

Publication:

García-Murria, M.J.*, **Duart, G**.*, Grau, B., Diaz-Beneitez, E., Rodríguez, D., Mingarro, I., Martínez-Gil, L., 2020. *Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control*

cellular apoptosis. **Nature Communications** 11, 6056. https://doi.org/10.1038/s41467-020-19881-9
*Equal contribution

Programmed cell death is a fundamental process in the development of multicellular organisms contributing to the balance among cell death, proliferation, and differentiation, that is crucial for tissue development and homeostasis (Kerr et al., 1972). Also, protection and defence against many disorders, including cancer and pathogen-related diseases, rely on apoptosis to eliminate the affected cells (Häcker, 2018; Hua et al., 2019).

One of the primary modulators of apoptosis is the B-cell lymphoma 2 (Bcl2) protein family (Kim et al., 2006). The proteins in this family can be divided in anti-apoptotic (e.g., Bcl2 and BclxL) (Boise et al., 1993), pro-apoptotic (e.g., Bax and Bak)(Oltvai et al., 1993), and BH3-only apoptosis activators (e.g., Bid and Bmf) (Wang et al., 1996). Most pro-apoptotic and anti-apoptotic proteins in this family share up to four main Bcl2 sequence homology domains, known as BH1, BH2, BH3, and BH4, while BH3-only members have solely the BH3 domain. In addition, many Bcl2 family members have a TMD in the carboxyl-terminal end that effectively allows for insertion of the protein into the target lipid bilayer (Delbridge et al., 2016).

Cellular Bcl2 (cBcl2) proteins can physically interact with each other, forming homo-and hetero-oligomers that are crucial for programmed cell death regulation (Cosentino and García-Sáez, 2017; Kelekar et al., 1997; Oltvai et al., 1993; O'Neill et al., 2006; Wang et al., 1996; Xie et al., 1998). To prevent the premature death of host cells, viruses have developed functional homologues of cBcl2, known as viral Bcl2 (vBcl2), as a strategy to modulate cell death (Kvansakul et al., 2017; Polčic et al., 2017). Although there is low sequence homology between vBcl2 and cBcl2, crystal structures reveal a structural homology in key domains (Galluzzi et al., 2008; Kvansakul and Hinds, 2013).

In this work, we firstly proved that vBcl2 contain a functional TMD in the Ct end, like their cellular counterparts. We selected six vBcl2 proteins from two distinct viral families (3 hepesviruses and 3 poxviruses). BHRF1 (*Human gammaherpesvirus 4 – Epstein–Barr virus*, HHV4) (Pearson et al., 1987), ORF16 (*Human gammaherpesvirus 8 – Kaposi's sarcoma–associated herpesvirus*, HHV8) (Cheng et al., 1997), ORF16 (Bovine gammaherpesvirus 4, BoHV4) (Bellows et al., 2000), F1L (*Vaccinia virus*, VacV) (Nichols et al., 2017; Wasilenko et al., 2003), M11L (*Myxoma virus*, MyxV) (Douglas et al., 2007; Nichols et al., 2017), and ORFV125 (*Orf virus*, OrfV) (Westphal et al., 2007). To avoid confusion, here we use the viral acronym to refer to the vBcl2 protein. *In silico* analyses suggested the presence of TMDs in the selected vBcl2. Then, we aimed to explore the membrane insertion capacity of the predicted segments using an *in vitro* assay based on the *E. coli* leader peptidase (Lep) described before (see *section* 3.*1*). Using this glycosylation-based system we determined that all the studied vBcl2 regions were efficiently inserted into microsome membranes.

After determining that the vBcl2s included a Ct end anchoring TMD, we sought to study whether their role went beyond that of a structural anchor. As interactions between cBcl2 TMDs have been reported in biological membranes (Andreu-Fernández et al., 2017), we decided to study the vBcl2 TMD capability of homo and hetero-oligomerization in biological membranes. For this purpose, we used two different systems of bimolecular complementation based on different reporters and in different model organisms. We used BiFC (bimolecular fluorescent complementation) approach (Kerppola, 2006) adapted for the study of intramembrane interactions in eukaryotic cells (Andreu-Fernández et al., 2017; Grau et al., 2017). This technique is based on a split venus fluorescent protein (VFP). Each of the two non-fluorescent fragments of the VFP is fused to one of the studied TMDs and expressed in eukaryotic cells. The two fragments of the VFP have no affinity for each other. The VFP will be reconstituted (and its fluorescence) only when TMD-TMD interaction is reported. The TMDs of HHV4, HHV8, VacV, MyxV, and OrfV showed a homo-interaction capability above the controls and similar to that observed with cBcl2 TMD.

However, BoHV did not show VFP-associated fluorescence significantly higher than the negative controls. Western blot analysis showed comparable expression levels of all chimeras.

To investigate the potential TMD-TMD heteromeric interactions between vBcl2 and cBcl2 we used the previously described BiFC approach. We investigated the potential TMD-TMD interactions between vBcl2 and the anti-apoptotic cellular Bcl2 and BclxL proteins. Interestingly, all viral TMDs included in the assay could interact with the TMD of Bcl2. However, although the majority of vBcl2 TMDs also interacted with BclxL TMD, BoHV and MyxV TMDs did not. We also analysed the interactions between vBcl2 TMDs and the TMDs of the cellular pro-apoptotic Bax and Bak proteins. In the absence of apoptotic stimulus and independent of any cytosolic (soluble) domain contacts, all three poxviral TMDs (VacV, MyxV, and OrfV) could interact with Bax and Bak TMDs. On the contrary, herpesviruses HHV4, HHV8, and BoHV TMDs showed no interaction with Bax or Bak TMDs. Finally, we included in the BiFC-based screening the TMDs from BH3-only apoptotic modulators Bik and Bmf (Andreu-Fernández et al., 2016). The TMDs of HHV4, HHV8, VacV, and OrfV could interact with the TMD of Bik. However, viral interactions with the TMD of Bmf were limited, only the HHV8 TMD interacts with Bmf TMD. Also, we further explored some of these interactions (e.g., HHV8-Bcl2 and MyxV-Bax) using the BLaTM assay (Schanzenbach et al., 2017) and computational modelling. BLaTM is a genetic tool designed to study TMD–TMD interactions in bacterial membranes. This tool is based on the use of a split β-Lactamase fused to the TMDs of study and expressed in *E. coli* cells. The TMD-TMD interaction aims the reconstitution of the β-Lactamase activity and is reported by ampicillin resistance. In this assay, the LD_{50} of the antibiotic served as an indicator of the interaction strength. Of note, bacteria can grow only in the presence of ampicillin when the β-lactamase is reconstituted in the periplasm. Therefore, the BLaTM assay also reports the insertion of the tested regions. With this system we double checked some of the previously studied interactions in a more quantitative manner. Overall, our findings point out an intricate network of interactions between the TMDs of viral and cellular origin.

Finally, we wanted to determine if the TMD-TMD interaction network described played a role in apoptosis modulation. To that end, we transfected HeLa cells with Bcl2, HHV8, or MyxV either with or without the TMD (FL and ΔTMD, respectively). We also included chimeras in which the TMD of each of the previously described proteins was replaced by the TMD of the non-apoptotic mitochondrial protein TOMM20 (Bcl2-T20, HHV8-T20, and MyxV-T20, respectively), which our earlier studies suggested could not interact with any viral or cellular Bcl2 TMD. Cells were transfected with the appropriate constructs and then treated with doxorubicin to induce apoptosis (Rooswinkel et al., 2014). As expected, the FL proteins prevented apoptosis. When the TMD was removed, however, none of the proteins retained their anti-apoptotic effect. Similarly, the chimeras carrying the TMD of T20 could not control the doxorubicin-induced apoptosis, although localizing in the same membranes as the FL protein. These experiments were also replicated using different apoptotic stimulus as viral-induced apoptosis or Bax-induced apoptosis (only for Bcl2 and MyxV; FL and T20 variants).

In summary, we have identified the Ct hydrophobic region of the vBcl2 as a genuine TMD capable of interacting with cBcl2 TMDs. We have also demonstrated that these intramembrane interactions are critical for viral cell fate control. This study provides a deeper understanding about how viruses control cellular death for their own benefit and contribute to our better understanding of how viruses interact with their hosts.

3.4. Chapter 4: Membrane protein interaction design

Publication:

Duart, G., Elazar, A., Weinsten, J, Gadea-Salom, L., Ortiz-Mateu, J., Fleishman, S., Mingarro, I., Martínez-Gil, L. *Computational design of BclxL inhibitors that target transmembrane domain interactions. Under revision.*

In this chapter, we aimed to explore the intramembrane protein–protein interactions (PPIs) of BclxL protein and their role in its anti-apoptotic function. We used the information obtained for computationally design an inhibitor capable of selectively sequester the TMD of BclxL to turn back apoptosis resistance.

As explained in *section 2.3*, the members of the Bcl2 protein family can interact with each other, forming homo- and hetero-oligomers (Andreu-Fernández et al., 2017; Cosentino and García-Sáez, 2017; Kelekar et al., 1997; Oltvai et al., 1993; Wang et al., 1996). These PPIs are part of an important regulatory network of mitochondrial outer membrane (MOM) permeabilization driving programmed cell death. In healthy cells, anti-apoptotic Bcl2 members inhibits activation of pro-apoptotic proteins through direct interaction or by sequestering BH3-only proteins (Kim et al., 2006). Upon an apoptotic stimulus, BH3-only and pro-apoptotic proteins are released and free to induce MOM permeabilization. Interactions among Bcl2 family members have been thought to occur through soluble domains, especially BH domains (Dadsena et al., 2021). However, recent findings demonstrate that their TMDs also participate in these PPIs (Andreu-Fernández et al., 2017; Lucendo et al., 2020) and, as proved in Chapter 3, these intramembrane interactions are crucial for apoptotic control.

Among the anti-apoptotic proteins, the Bcl-2–like protein 1, better known as BclxL, displays relevant functions in several forms of cancer. In melanoma, BclxL participates preventing cells from executing apoptosis, and inducing drug resistance, cell migration and invasion, and angiogenesis (Lucianò et al., 2021). Because of the relevance of BclxL in the progression of cancer, different strategies have been considered to inhibit it (Lucianò et al., 2021) but new and less adverse BclxL inhibitors are needed.

First, we sought to determine if the BclxL TMD can interact with other pro- and anti-apoptotic Bcl2 members. The potential intramembrane contacts were assessed using the genetic tool BLaTM (see *section 3.3*). Using this assay, we tested the homo-oligomerization of BclxL TMD and its putative hetero-oligomerization with the TMD of anti-apoptotic Bcl2, and

the TMDs of pro-apoptotic (Bax and Bak) and BH3-only (Bik and Bmf) Bcl2 members. According to our findings, BclxL TMD forms weak homo-oligomers. Furthermore, we identified TM hetero-oligomers with Bcl2, Bax, and Bak. To corroborate the interaction capabilities of the BclxL TMD in eukaryotic cells, we used the BiFC assay (Kerppola, 2006), adapted for the study of intramembrane interactions (Andreu-Fernández et al., 2017; Grau et al., 2017) (see *section 3.3*). Our results indicated that the BclxL TMD can homo-oligomerize, and hetero-oligomerize with the TMDs of Bcl2, Bak, and Bik in eukaryotic membranes.

Next, to investigate whether these TMD–TMD interactions are necessary for BclxL cellular apoptosis regulation, we transfected HeLa cells with full-length BclxL with or without the TMD (BclxL-FL and BclxL-ΔTMD, respectively). Additionally, we included a chimera in which the TMD of BclxL was replaced by the TMD of TOMM20 (BclxL-T20). Our results indicated that when doxorubicin is used as a cell death stimulus, BclxL requires the TMD to block apoptosis.

As BclxL localizes primarily in the mitochondria (Fang et al., 1994; González-García et al., 1994; Zamzami et al., 1998), and given the importance of TMDs for membrane protein sorting (Martínez-Gil et al., 2011), we explored how deletions or substitutions in the Ct hydrophobic region of BclxL affected cellular localization. To analyze the subcellular location, BclxL (BclxL-FL), the BclxL-T20, and BclxL-ΔTMD variants were expressed in HeLa cells alongside a fluorescent mitochondrial marker. The fluorescence micrographs revealed that both BclxL-FL and BclxL-T20 moieties were located at the mitochondria while BclxL-ΔTMD, as expected, showed a cytosolic distribution.

Once established that BclxL TMD–TMD interactions were crucial for the anti-apoptotic function, we aimed to design a specific inhibitor for these intramembrane PPIs. Inhibitor design started with the modeling of the BclxL TMD homo-interaction using TMHOP (Trans-membrane Homo Oligomer Predictor) (Weinstein et al., 2019). TMHOP uses Rosetta symmetric all-atom *ab initio* fold-and-dock simulations in an implicit

membrane environment to predict thousands of low-energy conformations based on the energy function that relies on the empirical measurement of amino acid insertion propensities (Elazar et al., 2016; Weinstein et al., 2019). Based on structural characteristics and associated Rosetta energy, we selected a TMHOP model that forms a tightly packed parallel dimer. The selected model was then fed into the FuncLib design algorithm (Khersonsky et al., 2018) to generate higher affinity binders that could serve as inhibitors. The FuncLib algorithm uses Rosetta design calculations to enumerate combinations of tolerated amino acid substitutions at specific positions. It then relaxes each combination using whole-protein minimization (based on the Rossetta membrane energy function) (Weinstein et al., 2019) and ranks these combinations by energy. Since single-span TMDs are known to self-associate (Grau et al., 2017), we designed sequences with minimal self-association potential by including the following rules: i) we intend to design a sequence able to interact with the BclxL wt TMD. For that, we used positive selection for a heterodimer (Non-symmetric FuncLib; $\Delta\Delta G<+1$ Rosetta energy units; R.e.u.). ii) we use negative selection for a new homodimer (symmetric FuncLib; $\Delta\Delta G>+5$ R.e.u.). As a consequence, the algorithm will only include diversification in the sequence, which will allow the new sequences to efficiently interact with the native BclxL TMD sequence but will not allow the inhibitor to interact with itself. This process resulted in three designed potential TMD inhibitor sequences (named D1, D2, and D3).

We verified that these designed segments could be inserted into ER-derived microsomes using an *in vitro* transcription/translation assay. Next, using BLaTM, we analyzed the interactions between the TMD of BclxL and the computationally designed inhibitors D1, D2, and D3. The results of these experiments revealed that D1 can efficiently bind to the BclxL TMD but does not form homo-oligomers, as we intended in our design. Of note, the interaction between the TMD of BclxL and D1 was stronger than the homo-oligomerization of the TMD of BclxL. Although D2 and D3 did not form homo-oligomers, they did not interact with the TMD of BclxL. Also, we investigated the specificity of the observed interaction by challenging D1 with the TMD of Bcl2, another anti-apoptotic protein. We detected no

interaction between D1 and the Bcl2 TMD. Thus, any effect of D1 on cell survival would most likely arise from its interaction with the BclxL TMD. Additionally, we used the BiFC assay to ensure that the interaction between D1 and the BclxL TMD was maintained in eukaryotic membranes. The results indicated that D1 could efficiently bind to the TMD of BclxL and that did not form homo-oligomers in eukaryotic cells.

To inhibit the anti-apoptotic effect of BclxL, the designed sequences must be in the same cellular compartment where BclxL is found. To test their location, we fused D1 and D2 sequences to the Ct of the enhanced green fluorescent protein eGFP (eGFP-D1 and eGFP-D2) and expressed these constructs in HeLa cells together with BclxL attached to the fluorescent protein mCherry (mCherry-BclxL). Next, we analyzed the subcellular distribution of these chimeras by confocal fluorescence microscopy. Both eGFP-D1 and eGFP-D2 showed a strong co-localization with mCherry-BclxL. To ensure that D1 and BclxL coexist in the same cellular compartment we performed a second localization assay based on organelle differential ultracentrifugation (Geladaki et al., 2019). We then analyzed the protein content in each fraction by sequential window acquisition of all theoretical mass spectra (SWATH-MS) (Rotello and Veenstra, 2021; Zhang et al., 2020). BclxL and eGFP-D1 had a similar distribution profile, suggesting a similar subcellular localization.

Finally, we tested the anti-apoptotic effect of D1 and D2. HeLa cells were transfected with BclxL alongside the eGFP-T20, eGFP-D1, eGFP-D2, or eGFP-xL chimeras. As a control, we used cells that did not receive BclxL or any of the chimeras and transfected them with an empty plasmid to keep the amount of transfected DNA constant across all samples. After transfection, cells were treated with doxorubicin to induce apoptosis. The cells that received eGFP-T20 or eGFP-xL plus BclxL could block doxorubicin-induced apoptosis. Remarkably, transfection of eGFP-D1 eliminated the anti-apoptotic effect of BclxL. D2 also reduced cell viability but less drastically than D1. Of note, no significant differences were found between the samples transfected with eGFP-D1 or eGFP-D2 (plus BclxL), or cells transfected with an empty plasmid treated in all cases with doxorubicin,

indicating that D1 and D2 are capable of inhibiting BclxL function. We also tested cell viability after the transfection, proving that neither D1 nor D2 was toxic to HeLa cells, a vital characteristic when designing non-toxic inhibitors.

In summary, these results have provided evidence of the importance of TMD–TMD interactions in apoptosis control, particularly in the case of BclxL. We successfully designed sequences capable of specifically inhibit the anti-apoptotic action of BclxL. Our work shows a path to design effective inhibitors based only on the sequences of the target receptor. The fact that two of three designs exhibited the desired hetero- and no homo-interactions highlights the accuracy of the TMHOP modelling strategy and the FuncLib design algorithm, which has already been applied to a wide range of soluble protein design tasks. This work significantly advances our understanding of sequence-specific recognition in membranes and opens the way for a new generation of anti-cancer drugs.

4. Publications

4.1. Chapter 1: Insertion
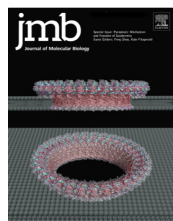
Intra-Helical Salt Bridge Contribution to Membrane Protein Insertion

Gerard Duart [1,†], **John Lamb** [2,†], **Juan Ortiz-Mateu** [1], **Arne Elofsson** [2,*] and **Ismael Mingarro** [1,*]

1 - Departament de Bioquímica i Biologia Molecular, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, E-46100 Burjassot, Spain
2 - Science for Life Laboratory and Department of Biochemistry and Biophysics, Stockholm University, 171 21 Solna, Sweden

Correspondence to Arne Elofsson and Ismael Mingarro: Ismael.Mingarro@uv.es *(I. Mingarro),* @GerardDuart 🐦 *(G. Duart),* @bio_lamb 🐦 *(J. Lamb),* @arneelof 🐦 *(A. Elofsson),* @IsmaelMingarro 🐦 *(I. Mingarro)*
https://doi.org/10.1016/j.jmb.2022.167467
Edited by William Clemons

Abstract

Salt bridges between negatively (D, E) and positively charged (K, R, H) amino acids play an important role in protein stabilization. This has a more prevalent effect in membrane proteins where polar amino acids are exposed to a hydrophobic environment. In transmembrane (TM) helices the presence of charged residues can hinder the insertion of the helices into the membrane. It is possible that the formation of salt bridges could decrease the cost of membrane integration. However, the presence of intra-helical salt bridges in TM domains and their effect on insertion has not been properly studied yet. In this work, we show that potentially salt-bridge forming pairs are statistically over-represented in TM-helices. We then selected some candidates to experimentally determine the contribution of these electrostatic interactions to the translocon-assisted membrane insertion process. Using both *in vitro* and *whole cell* systems, we confirm the presence of intra-helical salt bridges in TM segments during biogenesis and determined that they contribute ~0.5 kcal/mol to the apparent free energy of membrane insertion (ΔG_{app}). Our observations suggest that salt bridge interactions can be stabilized during translocon-mediated insertion and thus could be relevant to consider for the future development of membrane protein prediction software.

Introduction

Most integral membrane proteins have to insert their transmembrane (TM) segments into the lipid bilayer in a helical conformation and then acquire a defined three-dimensional structure by packaging their helices[1]. α-Helical TM segments are largely composed of apolar residues because of the hydrophobic nature of the membrane. Nevertheless, in some cases, it is necessary for the protein activity to include polar amino acids in a TM region.[2] This fact is sometimes not contemplated in modern membrane topology prediction tools,[3,4]

in which the presence of charged amino acids in a sequence automatically supposes a penalty increase in the predicted free energy (ΔG_{pred}) of insertion. The presence of polar amino acids in TM regions is more frequent than what would be expected according to the hydrophobic nature of the environment,[5] especially when these are in pairs on the same face of an α-helix.

Salt bridges are electrostatic interactions between negatively (D, E) and positively charged (K, R, H) amino acids that play an important role in protein stabilization.[6] Many studies have shown that pairs of charged residues that form potential

salt-bridges stabilize soluble α-helices.[7] Salt bridges have a prominent role in the folding of globular proteins and, despite their low occurrence in TM domains, they might contribute to membrane protein stability. This contribution is especially important in membrane protein biogenesis, as salt bridges might lower the polarity of charged residues in the hydrophobic environment of the Sec-translocon,[8] suggesting that salt bridges could aid the insertion of marginally hydrophobic TM α-helices.[5,9]

To investigate the potential formation of intra-helical salt bridges in TM α-helices, we analyzed the composition of the TM domains from membrane proteins of known structures. This analysis showed that charged residue pairing is more prevalent than expected for pairs located on the same face of TM α-helices. Likely, salt bridge formation on the same face of α-helices reduces the unfavorable energetics of inserting charged residues into the membrane.[10] Then, we generate a list of potential candidates for further *in vitro* and *whole cell* experiments.

We found that the apparent free energy (ΔG_{app}) of membrane insertion through the translocon machinery can be decreased by ∼0.5 kcal/mol by position-specific paired charges interaction. These findings lead to a better understanding of the insertion mechanism of TM helices and can in the future be used to improve topology prediction tools.

Results

Paired charges interactions in model transmembrane helices

To test the contribution of potential salt bridges to the translocon-mediated membrane insertion, we used as a vehicle the leader peptidase (Lep) protein from *Escherichia coli* (Figure 1(A)). The Lep protein consists of two TM segments (H1 and H2) connected by a cytosolic loop (P1) and a large C-terminal (P2) domain, which inserts into endoplasmic reticulum (ER)-derived rough microsomes with both termini located in the microsomes lumen. The designed TM segments were inserted into the luminal P2 domain and flanked by two acceptor sites (G1 and G2) for N-linked glycosylation. Glycosylation occurs exclusively in the lumen of the ER (or microsomes) because of the location of the oligosaccharyltransferase (OST) active site (a translocon-associated enzyme responsible for the oligosaccharide transfer).[11] In this case, the engineered glycosylation sites can be used as membrane insertion reporters because G1 will always be glycosylated due to its native luminal localization, but G2 will be glycosylated only upon translocation of the analyzed sequence across the microsomal membrane. A singly glycosylated construct in which a tested sequence is inserted into the membrane has a molecular mass ∼2.5 kDa

higher than the molecular mass of Lep molecule expressed in the absence of microsomes; the molecular mass shifts by ∼5 kDa upon double gly-cosylation, which facilitates its identification by gel electrophoresis when expressed in the presence of [^{35}S-labeled] amino acids. The degree of membrane insertion is quantified by analyzing the fractions of singly glycosylated (i.e., membrane inserted) and doubly glycosylated (i.e., non-inserted) molecules, which can be expressed as an experimental apparent free energy of membrane insertion, ΔG_{exp} (see Materials and Methods).[12–14]

We first compared the effects of oppositely charged Lys and Asp residues on the insertion of a 19-residue-long hydrophobic stretch (L4/A15 scaffold, 4 leucine and 15 alanine), which was designed to insert stably into the RM membranes[12] (Figure 1(B)) and "insulated" from the surrounding sequence by N- and C-terminal GGPG- and -GPGG tetrapeptides. Single Lys and Asp residues were placed in positions 8 and 12 respectively, and pairs of Lys-Asp residues were designed to cover positions 7–12 (that is, more than one helical turn). When pairs of charged residues are present, our results showed a tendency to insert more efficiently when oppositely paired charges were placed in positions (i, $i + 1$; i, $i + 3$; i, $i + 4$) that are permissive with salt bridge formation (Figure 1(C)), an effect not observed in the predictions (Figure 1(B)). Similar results were obtained on a different Leu/Ala background with a slightly higher insertion efficiency (L5/A14, 5 leucine and 14 alanine), those mutants harboring paired charges compatible with salt bridge formation (i.e., i, $i + 3$; i, $i + 4$) insert more efficiently than the non-compatible one i, $i + 5$ (Figure S1). Being the insertion of charged residues a thermodynamically inconvenient phenomenon within the membrane environment, it is expected that the sequence context and the amino acid composition of the TM helix would be determinant for salt bridge formation. Accordingly, we scrutinized a large dataset of membrane proteins of known three-dimensional structures to focus on potential salt bridges present within native TM segments.

Paired charges in transmembrane helices

We created two datasets, the TM dataset with α-helical membrane proteins of known structure from the PDBTM-dataset[15], and the GLOB dataset with globular α-helical proteins selected from the SCOP-database,[16,17] see methods for the full creation steps. Table 1 shows a breakdown of α-helices, charged residues and salt bridges in the two datasets. Long α-helices (≥17 residues) form a larger proportion in the TM dataset than in the globular helices (GLOB) dataset, as TM helices need to span the hydrophobic core of the lipid bilayer of thickness ∼30 Å.

Our TM dataset showed the same distribution of polar charges as previous studies,[18] with about

10% of polar residues in the core membrane regions (see Table 1), with under 4% being charged. Over half of the charged residues could form pairs with other charged residues at intervals of i, $i + 1$; i, $i + 3$ and i, $i + 4$. In contrast, the GLOB dataset had a much higher proportion of charged polar residues. The GLOB dataset also contained about 15 times as many paired charges relative to its size compared to the TM dataset, again indicating that charged residues are more common in globular than in TM helices.

The two datasets, TM and GLOB, were extended with homologous sequences identified by searching with jackhammer against UniProt. All sequences with an E-value lower than 10^{-3} were included. These datasets are named TM-MSA and GLOB-MSA. Using data in the TM-MSA dataset to produce log odds ratios, we identified periodicity patterns of paired charge residues that are more common than what would have been expected from the underlying amino acid composition (see Figure 2 and Table 2). We found that polar residues at pairs i, $i + 3$; i, $i + 4$ and i, $i + 7$ are significantly enriched (Figure 2). This feature is strengthened when we examined the same plot for the GLOB-MSA dataset, where these patterns were not observed (Figure S2). This was also clear when statistical significance was taken into account, see Figures S3 and S4.

Structural analysis of charged residues in transmembrane helices

As expected, globular proteins contain more local salt-bridges in α-helices, 13% *vs* 5% in membrane proteins, see Table 1. This does conform to our
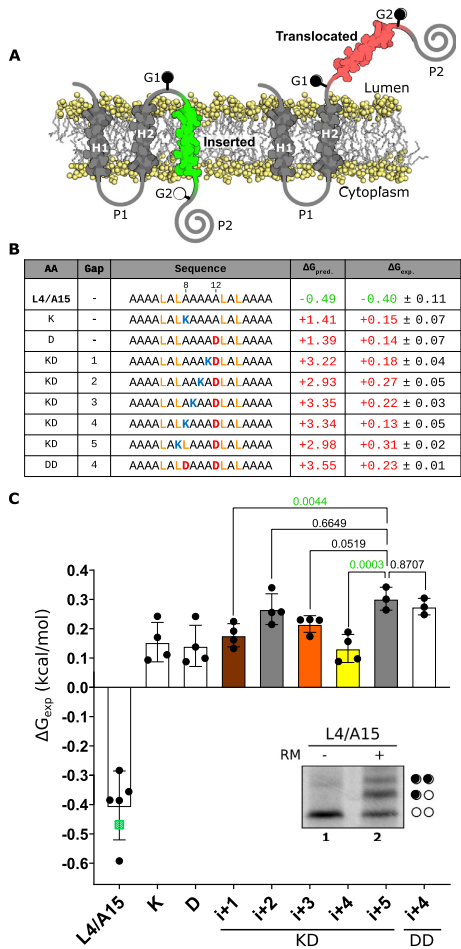
Figure 1. Effects on membrane insertion of single or pairs of Asp and Lys residues in a model TM segment. (A) Schematic representation of the leader peptidase (Lep) model protein. G1 and G2 denote artificial glycosylation acceptor sites. The sequence under investigation was introduced in the P2 region after H2. Recognition of the tested sequence as a TM by the translocon machinery (highlighted in green) results in the modification of the G1 site but not G2. The Lep chimera will be double glycosylated if the sequence being tested is not recognized as a TMD and thus translocated into the microsomes lumen (shown in red). (B) The tested sequences from L4/A15 model TM (including the charged residues, bold), the gap distance, and the predicted ΔG (ΔG_{pred}) values in kcal/mol are shown. Amino acids with positive and negative charge are highlighted in blue (K) and red (D) respectively. (C) Experimental ΔG (ΔG_{exp}) in kcal/mol of each tested sequence in the Lep-based microsomal assay. The mean and standard deviation of at least 3 independent experiments is represented (*n* values: 4 [from 1 to 7] and 3 [8 and 9]). The individual value of each experiment is represented by a solid dot, *p*-values (ordinary one-way ANOVA test with Dunnett correction) are indicated above the corresponding bars with values <0.005 highlighted in green. In addition, a green square represents the experimental ΔG value for the L4/A15 sequence from an earlier study.[12] The wt and single mutants are shown in white bars. Charges at compatible distances with salt bridge formation (i, $i + 1$; i, $i + 3$; and i, $i + 4$) are shown in brown, orange and yellow, respectively. Not compatible distances with salt bridge formation (i, $i + 2$; and i, $i + 5$) are shown in dark gray. The inset shows a representative SDS-PAGE gel for L4/A15 construct. The construct was expressed in rabbit reticulocyte lysed in the presence (+RM) or absence (−RM) of rough microsomes. Bands of non-glycosylated proteins are indicated by a white dot; mono and double glycosylated proteins are indicated by one and two black dots, respectively.

Table 1 About 10% of the core residues in the TM dataset are polar and just under 4% are charged residues. This is significantly less than 37.5% and 25.5% respectively in the globular dataset. In the transmembrane dataset, just over half of the charged core residues form a charged pair at distance 1, 3 or 4. *In the case of the globular set, there are more potential charged pairs than charged core residues as multiple residues are counted more than once as they form more than one potential pairing with separation of 1, 3 or 4 steps. It is clearly seen that although the TM dataset contains more helices per protein and has a higher proportion of long (≥17 residues) helices it contains significantly fewer charged pairs.

Dataset	TM	TM-MSA	GLOB	GLOB-MSA
Charge statistics				
Core residues	40,116	3,990,464	74,299	7,388,277
Polar core residues	4080 (10.2%)	490,861 (12.3%)	27,851 (37.5%)	2,696,564 (36.5%)
Charged core residues	1467 (3.7%)	204,544 (5.1%)	18,959 (25.5%)	1,819,047 (24.6%)
Pairs of charged core residues (+1, +3 or +4)	825	88,380	23755*	2271405*
# of same charged pairs	373	45,220	10,835	1,117,270
# of oppositely charged pairs	452	43,160	12,920	1,154,135
Charge Pairs statistics				
Proteins	925	130,934	2475	363,610
Total α-helices	5435	481,729	23,991	2,118,158
α-helices ≥ 17	3755	306,632	5895	485,797
Proteins with charged pairs	204	25,187	1780	193,534
Helices with charged pairs	265	34,029	4251	344,142
Proteins with any salt bridge	241	–	751	–
Proteins with local salt bridge in helix	56	–	335	–
Total number of local salt bridges in helices	59	–	577	–

current understanding of soluble versus TM α-helices and their different environments. However, as seen in Figure 2, charged pairs of amino acids are especially prevalent at positions i, $i + 1$; i, $i + 3$ and i, $i + 4$. Oppositely charged residues stand out, especially Glu-Arg at i, $i + 1$, Glu-Lys at i, $i + 3$ and Asp-Lys at i, $i + 4$. Also, several same charged pairings at i, $i + 3$ and i, $i + 6$ are more frequent than expected. Other known structural features can also be hinted at, including aromatic ring stacking by His-Trp pair[19] at i, $i + 6$, and contacting with prosthetic groups by His-His pair at i, $i + 7$.[18]

Paired charges placed at i, $i + 1$; i, $i + 3$ and i, $i + 4$ could potentially form salt bridges as they are all on the same relative face of the α-helix and are close enough in vertical separation on the helix (see Figure 3, top). Although oppositely paired charges at i, $i + 7$ are also on the same face of the α-helix, unless the α-helix has a bend, both residues are too far separated to form a salt bridge. This was clearly seen in Figure 3 where the TM dataset was used. In both absolute count and log odds ratios (Figure 3(A)) it is clear that residues at i, $i + 1$; i, $i + 3$ and i, $i + 4$ are by far the most common and over-represented pairings. Figure 3 (A) also shows that oppositely and same charged pairs have about the same over-representation at i, $i + 3$, whereas oppositely charged pairs are stronger than same paired charges at i, $i + 1$ and i, $i + 4$, both within salt bridge range, and same paired charges are more prevalent at positions i, $i + 7$ and i, $i + 8$, indicating that other effects than salt-bridges also play a role.

When structure-observed salt bridges in the different positions were compared to the oppositely paired charges a clear image aroused, see Figure 3(B). Even though there are more oppositely charged pairs at position i, $i + 1$ than in positions i, $i + 3$ with both log odds ratios over 1.0 (Figure 3(A)), only about 15% of the oppositely paired charges at i, $i + 1$ form salt bridge (Figure 3(B)). This contrasts with i, $i + 3$ where almost 40% of the pairs form salt bridges, and just under 25% at i, $i + 4$ (Figure 3(B)).

Selection of natural salt bridges from membrane protein structures

Each of the 8687 proteins in a redundant (TM-Red) dataset (see Materials and Methods) was scanned for oppositely paired charges in positions i, $i + 1$; i, $i + 3$ and i, $i + 4$ within any TM segments core region. For each of these, we only kept proteins with at least one TM segment that contains such a pair and where this pair was within salt bridging distance.

To select potential candidates, we also examined the ΔG_{pred} values and choose helices with a value above one (>+1 kcal/mol). This stricter definition results in a set of 426 candidates with a total of 431 salt bridges and a wide range of estimated ΔG_{pred} values (Figure S5). As shown in Figure S5, most TM segments with paired charges exhibit a surprisingly high ΔG_{pred} value that in normal circumstances would not insert into the membrane. Next, we selected TM7 (helix G) from halorhodopsin protein (PDB ID: 3QBG) with an
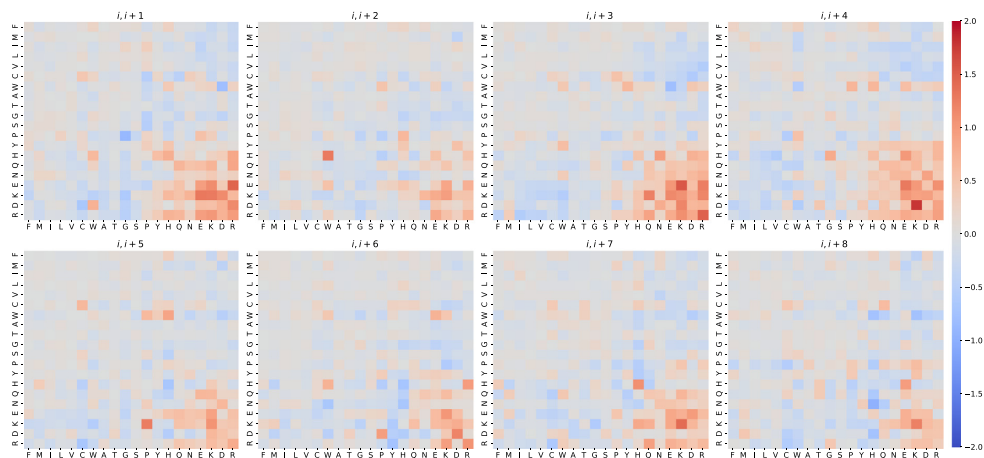
Figure 2. Log odds ratios of each pair of amino acids for '*i, i* + 1' through '*i, i* + 8' for the TM-MSA data. The rows on the y-axis indicate the first amino acid in the pair and the columns on the x-axis the second. The residues are ordered by hydrophobicity according to the Engelman order.[57] See Figure S1 for the equivalent of the globular dataset. S2 and S3 show the same plots masked for statistical significance.

estimated ΔG_{pred} value above +1.7 kcal/mol, and helix A from calcium ATPase (PDB ID: 1SU4) with a higher estimated ΔG_{pred} value (above +4.1 kcal/mol) for further studies. See the GitHub repository (https://github.com/ElofssonLab/salt_bridges) for the full TM helices list.

Intra-helical salt bridge stabilizes the insertion of helix G from halorhodopsin

Halorhodopsin (hR) from *Natronomonas pharaonis* (3QBG) is a protein made up of seven TM helices (helix A through G) and a retinal chromophore that is bound via a protonated Schiff base to the ε-amino group of a lysine (K258) located roughly in the middle of helix G.[20] *In silico*

analysis of 3QBG structure, the anion-free form of the protein, showed a paired charges (Asp-Lys), involving K258 and D254 residues from helix G (Figure 4). The distance between the anionic carboxylate ($RCOO^-$) from the Asp residue and the cationic ammonium (RNH_3^+) in the lysine residue, in the crystal structure, was about 3.5 Å, a permissive distance for a salt bridge formation (Figure 4 (G)), which has been established as being lower than 4 Å.[21] It is worth to mention that Asp-Lys pairs at *i, i* + 4 position are the most frequently found in the TM dataset (Figure 2). To get insights into this interaction, we designed three mutants that were supposed to perturb the salt bridge interaction in different ways: the K258D mutant place two charged residues with the same polarity at positions *i, i* + 4;

Table 2 Log odds ratios for all charged pairs, same charged pairs and oppositely charged pairs with calculated errors and multiple hypotheses corrected p-values. It is clear that charged pairs occur more often than predicted, evidenced by the positive log odds ratios in all cases. It is clear that positions +1, +3 and +4 are the most prevalent pairings, with the bolded values highlighting log odds ratio above 0.8. It is also clear that oppositely charged residues which have the potential to form salt bridges are prevalent in all three positions whereas the same charge is mainly prevalent in position 3. Most likely these same charges facing the same face of the helix are involved in functions such as ion transport.

| Spacing	All Log odds			Same charged Log odds			Oppositely charged Log odds		
	Ratio	*Error*	*p-value*	*Ratio*	*Error*	*p-value*	*Ratio*	*Error*	*p-value*
+1	*0.879*	*0.025*	*1.31e^{-257}*	*0.548*	*0.042*	*4.03e^{-41}*	*1.107*	*0.032*	*1.65e^{-256}*
+2	*0.252*	*0.037*	*2.02e^{-08}*	*0.185*	*0.054*	*1.88e^{-00}*	*0.316*	*0.050*	*1.09e^{-06}*
+3	*1.073*	*0.026*	*<0.00e^{-300}*	*1.118*	*0.035*	*1.15e^{-214}*	*1.026*	*0.037*	*5.78e^{-165}*
+4	*0.965*	*0.028*	*1.14e^{-253}*	*0.696*	*0.046*	*2.88e^{-49}*	*1.177*	*0.036*	*2.74e^{-233}*
+5	*0.517*	*0.037*	*4.88e^{-42}*	*0.390*	*0.055*	*4.15e^{-09}*	*0.630*	*0.049*	*1.21e^{-34}*
+6	*0.540*	*0.038*	*1.66e^{-43}*	*0.545*	*0.053*	*2.33e^{-21}*	*0.536*	*0.053*	*2.10e^{-20}*
+7	*0.615*	*0.038*	*1.21e^{-55}*	*0.764*	*0.050*	*7.78e^{-50}*	*0.439*	*0.058*	*1.97e^{-10}*
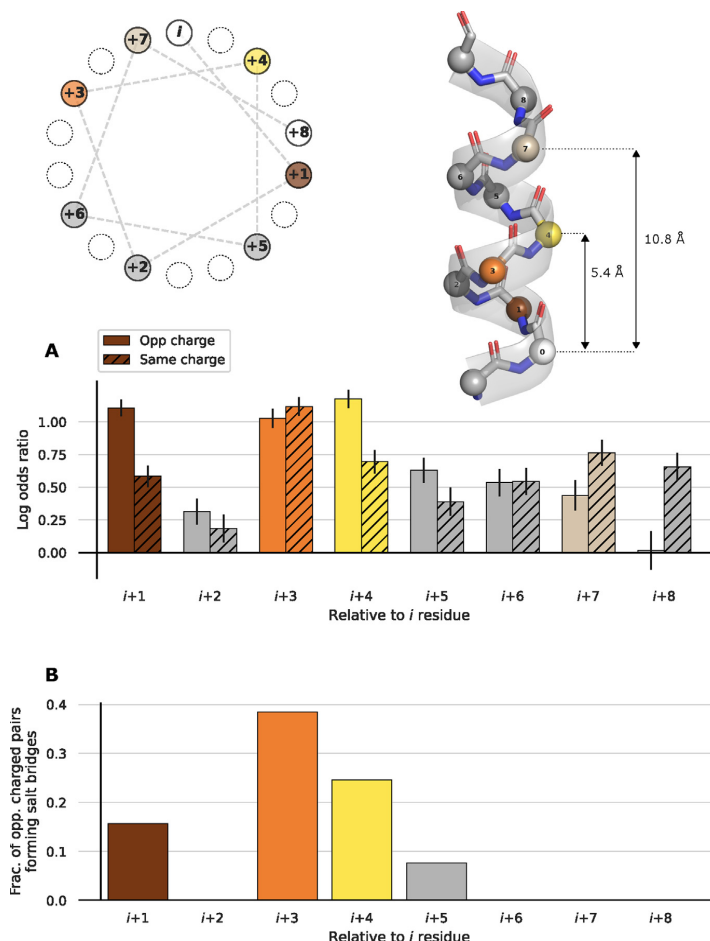
Figure 3. Charge pairs in TM helical sequences and structures. Helical wheel projection and lateral views of an α-helix are shown on top. The initial position i and the following 8 residues are numbered. Residues in positions $i + 3$ (orange), $i + 4$ (yellow), and $i + 7$ (light brown) are mainly on the same face of the helix, but $i + 7$ is placed too far for a salt bridge interaction. (A) The log odds ratios of charged pairs for '$i, i + 1$' through '$i, i + 8$' in the TM dataset. Plain filled bars refer to oppositely charged pairs and the forward slash are the and same charged pairs, all with error bars. (B) Fraction of oppositely charged pairs that form local salt bridges. The small bump at $i + 5$ are the two proteins 6CC4 (helix A) and 5MG3 (helix E), which both exhibit a bend in the α-helix due to the presence of glycine residues (see Figure S6).

the K258A mutant replaces one of the charged residues by a non-polar amino acid, and the K258Y/Y259K double mutant places the two charges at a non-permissive position for salt bridge formation (i, $i + 5$; Figure 4(G)), while keeping the same amino acid composition (Figure 4(A)).

Halorhodopsin is a trimeric protein in which the helix G is neither exposed at the monomer–monomer interface nor oriented to the inner part of the trimeric structure (Figure 4, panels (B–D)). The paired charges found in helix G are oriented toward the core of the 'globular' structure in each monomer. When constructs harboring helix G wild type sequence were translated *in vitro* in the presence of RMs singly-glycosylated (reporting insertion) forms were found (Figure 4(E), lane 2), despite its positive ΔG_{pred} value (Figure 4(A)). The nature of the higher molecular weight polypeptide species was analyzed by endoglycosidase H (EndoH) treatment, a highly specific enzyme that cleaves N-linked oligosaccharides. Treatment with EndoH of the
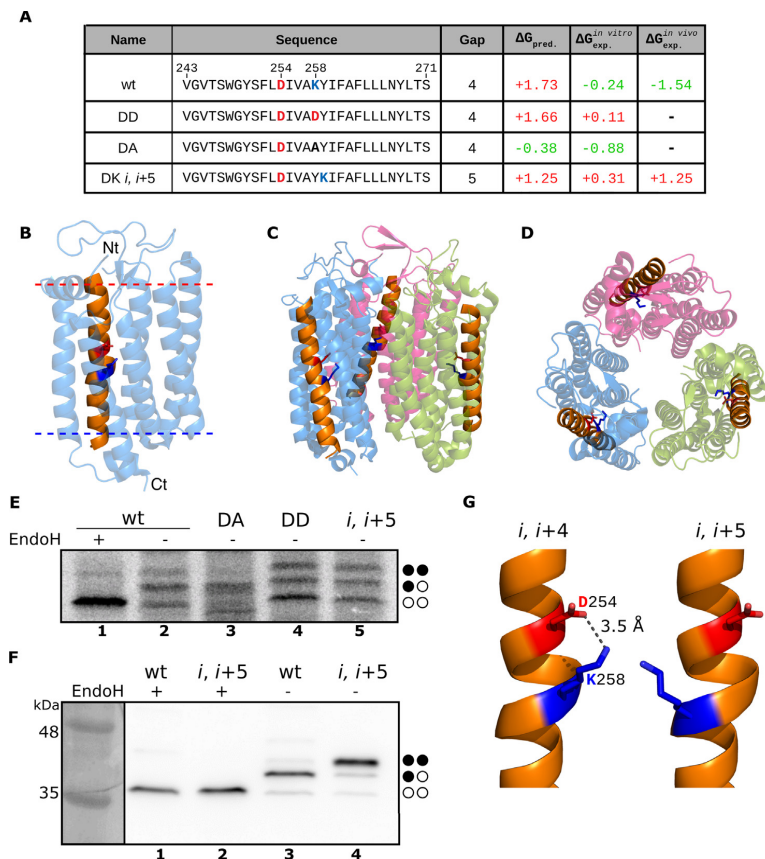
Figure 4. Insertion of halorhodopsin helix G from *Natronomonas pharaonis* (3QBG) into microsomal and cellular membranes. (A) Tested sequences from 3QBG including the gap distance, and the predicted (ΔG_{pred}) and experimental (*in vitro* $\Delta G_{exp}^{in\ vitro}$ and *whole cell* $\Delta G_{exp}^{in\ vivo}$, respectively) ΔG values in kcal/mol are shown. Amino acids with a positive or negative charge are highlighted in blue (K) and red (D), respectively. Green numbers indicate negative ΔG (insertion) values, while red numbers denote ΔG values above 0 (translocation). (B) Frontal view of 3QBG monomer structure. The helix G is highlighted in orange with the D254 and K258 shown in sticks colored red and blue respectively. The membrane position is indicated by a red (outer) and blue (inner) discontinuous line, according to OPM dataset. Lateral (C) and upper (D) views of the 3QBG trimeric structure. The helix G is highlighted in orange with the D and K shown in sticks colored red and blue, respectively. The different monomers are shown in transparent blue, pink and green. Representative examples ($n = 3$) of DNA construct translations in the presence of ER-derived microsomes (E) and Western blots ($n = 3$) of *whole cell* protein expression in HEK-293T cells (F) in the presence (+) or absence (−) of Endoglycosidase H (EndoH), a glycan-removing enzyme. The absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots. (G) Zoom view centered on the salt bridge between D254 and K257 at $i, i + 4$ (left) and $i, i + 5$ (right) gaps. D and K residues are shown in sticks colored red and blue, respectively, while the dashed line indicates the $RCOO^-$ to RNH_3^+ distance.

samples eliminated higher molecular mass bands (Figure 4(E), lane 1), confirming the sugar source of the retarded electrophoretic mobility bands and suggesting helix G insertion into the microsomal membrane. However, locating the Asp-Lys pair at $i, i + 5$ (Figure 4(G)), strongly reduced the experimental insertion efficiency (Figure 4(E),

lane 5). Replacing the positively charged lysine residue by a negatively charged aspartic acid residue (K258D) rendered similarly low levels of insertion efficiency (Figure 4(E), lane 4), and replacement of the ionizable lysine residue by the aliphatic alanine increased the insertion efficiency (Figure 4(E), lane 3).

The results of the Lep-based glycosylation assay indicated that wild type (wt) and K258A mutant are inserted properly into the microsomal membrane (ΔG_{exp} values −0.24 and −0.88 kcal/mol, respectively), but when the salt bridge is disrupted, either by having two charged amino acids with the same polarity (K258D) or by placing oppositely charged residues at a non-permissive distance (i, $i + 5$), the translocation of the segment increases substantially. It should be mentioned that the K258Y/Y259K double mutant has the same amino acid composition than the original helix G, but insertion efficiency is decreased (ΔG_{exp} = +0.31 kcal/mol). Together these results

show that the electrostatic interaction (salt bridge) between Asp and Lys residues in the center of the helix G from 3QBG is essential for its proper insertion into the microsomal membrane. The salt bridge contributes approximately ∼0.5 kcal/mol to the apparent experimental free energy of microsomal membrane insertion, as this is the difference found between the ΔG_{exp} values for the wt and i, $i + 5$ mutant.

Next, to ensure that the *in vitro* results are relevant to the *whole cell* situation, wt and i, $i + 5$ constructs were also expressed in HEK-293T cells. To this end, a *c-myc* tag was engineered at the C terminus of the Lep chimera to allow

Name	Sequence	Gap	$\Delta G_{pred.}$	$\Delta G_{exp.}^{in\ vitro}$	$\Delta G_{exp.}^{in\ vivo}$
L (L49-F78)	LWELVIEQFEDLLVRILLLAACISFVLAWF	4	+1.06	-1.05	-2.06
L i, i+5 (L49-F78)	LWELVIEQFEDLLVIRLLLAACISFVLAWF	5	+0.44	-0.34	-1.05
S (L49-A69)	LWELVIEQFEDLLVRILLLAA	4	+4.12	+0.22	-
S i, i+5 (L49-A69)	LWELVIEQFEDLLVIRLLLAA	5	+3.46	+0.06	-
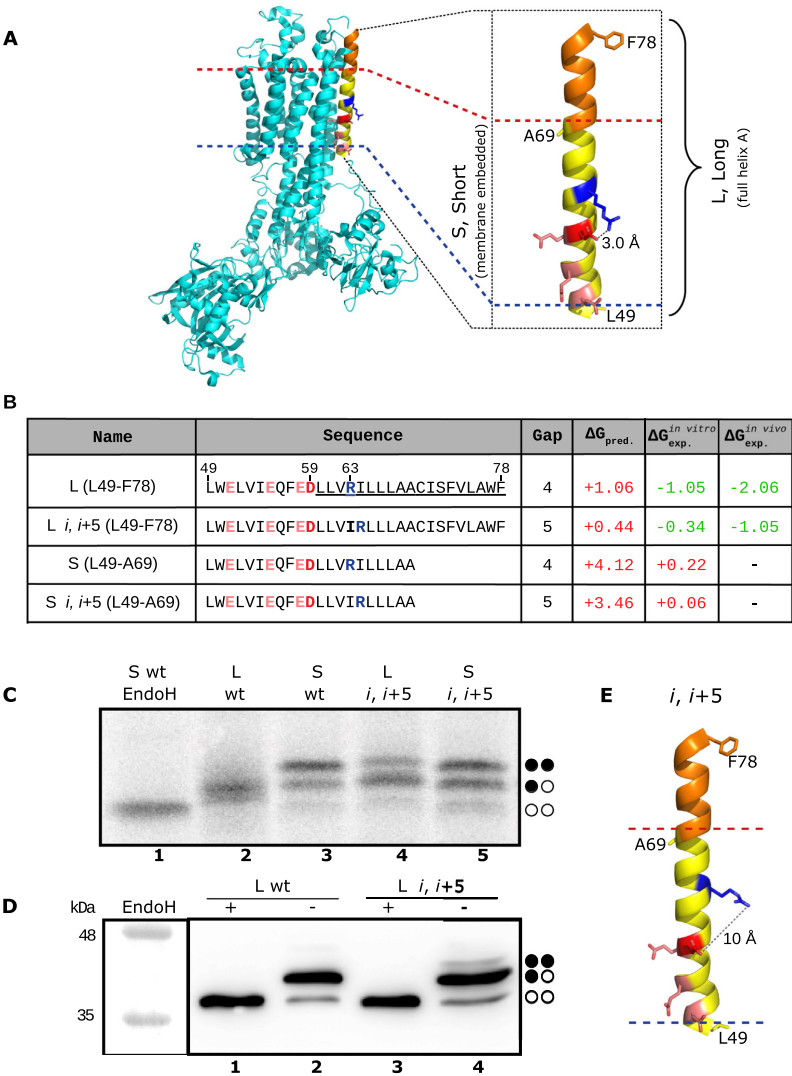
immune-detection of our constructs in cell extracts. As shown in Figure 4(F), transfected cells with the chimera containing helix G wt sequence rendered singly glycosylated molecules, indicating membrane insertion. In contrast, cells transfected with the construct harboring *i, i + 5* sequence rendered almost exclusively doubly glycosylated forms, as proved by EndoH treatment (Figure 4 (F), lane 2), suggesting membrane translocation. These results emphasized the relevance of salt bridge interactions in translocon-mediated TM insertion, especially in a cellular environment.

Salt bridge contribution to the insertion of a heavily charged helix

In order to examine salt bridge interactions in a more hydrophilic TM helix (Figure S5), we focus on helix A from the sarcoplasmic/ER calcium ATPase 1.[22] Calcium ATPase (PDB ID: 1SU4, *Oryctolagus cuniculus*) is a member of the P-type ATPases that transport ions across the membrane against a concentration gradient involving a bundle of 10 α-helices (helix A through J).[22] *In silico* analysis of the protein structure found an Asp-Arg (DR) pair (D59-R63) in the center of helix A. This helix extends beyond the membrane (Figure 5(A)) and have some particularities. On one hand, in the structure the N-terminal membrane-embedded region includes several charged amino acids (Glu51, Glu55, Glu58, Asp59 and Arg63), likely some of them involved in the Ca^{2+} transport across the membrane. Therefore, the ΔG_{pred} value for this segment (L49–A69) is remarkably high (+4.12 kcal/-mol, Figure 5(B)). On the other hand, the C-terminal region of this helix contains a mostly non-polar amino acid stretch that is compatible with the hydrophobicity of the membrane core. It has been previously shown that the position in the mem-

brane of TM helices in protein folded structures does not always correspond to regions being recognized by the translocon.[23] Instead, after translocon-mediated insertion of the more hydrophobic region, repositioning of TM helices relative to the lipid bilayer provides a way for non-hydrophobic polypeptide segments to become buried within the membrane. The sequence of helix A suggested the possibility that initial insertion of the hydrophobic region can be followed by subsequent repositioning of the charged region into the membrane hydrophobic core (Figure 5(A)). Supporting this the ΔG Predictor server (https://dgpred.cbr.su.se/) identified the C terminus, L60-F78 hydrophobic region (Figure 5(B)) and not the charged region (L49–A69) found in the protein structure.

The distance between the anionic carboxylate ($RCOO^-$) from D59 and the cationic guanidinium ($RC(NH_2)_2^+$) from R63 was about 3.0 Å in the crystal structure (Figure 5(A)), clearly within the permissive range for salt bridge formation. To investigate the contribution of this potential salt bridge interaction in the translocon-mediated membrane insertion of this region, we worked with two different scaffold sequences: the full helix A involving the residues 49–78 (Long, L); and a shorter membrane-embedded version including residues 49–69 (Short, S) as found in the solved structure. We also challenged the D59-R63 paired charges interaction in both sequences by increasing the separation between the ionizable residues from the native *i, i + 4* to non-permissive *i, i + 5*, while maintaining amino acid composition (R63I/I64R double mutant).

In vitro transcription/translation of these DNA sequences in the presence of microsomes rendered singly glycosylated molecules for the construct containing full-length helix A (Figure 5 (Clane 2). In contrast, when only the membrane-

---

Figure 5. Insertion of Calcium ATPase (1SU4) helix A into microsomal and cellular membranes. (A) Lateral view of 1SU4 structure. Zoom view of the A helix (right panel). The membrane-embedded region of helix A is highlighted in yellow. Charged amino acids are shown as sticks in blue (R), red (D) and pink (E), respectively. L49, A69 and F78 are also shown as sticks to define helix's subdomains. The membrane location is indicated by red (outer) and blue (inner) discontinuous lines according to OPM dataset[58] and the distance between the R and D charges is indicated in Å. (B) Helix A-derived sequences from 1SU4 including the gap between charged residues, and the predicted (ΔG_{pred}) and experimental (*in vitro* $\Delta G_{exp}^{in\ vitro}$ and *whole cell* $\Delta G_{exp}^{in\ vivo}$, respectively) ΔG values in kcal/mol are shown. Amino acids with a positive charge are highlighted in blue (K) while negatively charged are marked in red (D) and pink (E). The residues predicted as TM by the ΔG Prediction server are underlined. Green numbers indicate negative ΔG (insertion) values while red numbers denote ΔG values above 0 (translocation). Representative examples (*n* = 3) of *in vitro* protein expression in the presence of ER-derived microsomes (**C**) and Western blots (*n* = 3) of *whole cell* protein expression in HEK-293T cells (**D**) in the presence (+) or absence (−) of Endoglycosidase H (EndoH), a glycan-removing enzyme. The absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots. **E** 1SU4 helix A *i, i + 5* mutant. The membrane-embedded region of helix A is highlighted in yellow. Charged residues are shown as sticks in blue (R), red (D) and pink (E). L49, A69 and F78 are also shown as sticks to define helix's subdomains. The membrane is indicated by red (outer) and blue (inner) discontinuous lines as in (A), and the dashed line indicates the $RCOO^-$ and $RC(NH_2)_2^+$ distance.

embedded sequence (L49-A69) was included, the construct was mainly doubly glycosylated (Figure 5(C), lane 3), suggesting a reposition as described above. When the paired charges residues were placed at *i*, *i* + 5 (Figure 5(E)) the insertion efficiency was reduced (Figure 5(C), lane 4), with a ΔG_{app} decrease of $\sim$0.7 kcal/mol relative to the wild-type sequence (Figure 5(C), lane 2). As expected, this effect was not observed when the same mutations were grafted on the membrane-embedded (S) sequence (Figure 5(C), lane 5).

Next, we analyzed the salt bridge interaction in HEK-293T cells to study the translocon performance in cells. When cell cultures were transfected with a Lep-derived chimera containing the helix A wild type sequence only singly glycosylated molecules were observed (Figure 5 (D), lanes 1 and 2). However, a construct

harboring double mutant (R63I/I64R; *i*, *i* + 5) sequence showed doubly glycosylated molecules to a measurable extent (Figure 5(D), lanes 3 and 4), indicating a lower insertion efficiency.

Effect of salt bridge formation in the absence of previous TM regions

To investigate the effect of salt bridge formation in translocon-mediated membrane insertion in the absence of preceding (N-terminal) TM regions we used a different glycosylation-based reporter system in which the TM sequences of interest (bR helix G and ATPase helix A) were connected to the well-folded constant domain of the antibody λ light chain, C_L.[24] The C-terminal of the TM sequence carries an Asn-Val-Thr glycosylation site (G2), and we engineered an extra glycosylation site (G1) within the C_L sequence (Figure 6(A), top). As in the case of the Lep system, G1 will always be glycosylated due to the native translocation of the λ light chain, but G2 will be glycosylated only upon translocation of the analyzed sequence across the ER membrane (Figure 6(A), bottom). When constructs harboring either hR helix G or ATPase helix A wild type sequences were transfected into Hek293T cells, singly glycosylated (reporting insertion) forms were predominantly found (Figures 6(B), lane 1 and 6(C), lane 2, respectively), discarding any potential contribution to membrane insertion of preceding (N-terminal) TM segments. On the contrary,

Figure 6. Salt bridge effect on the insertion in the absence of preceding TM segments. (A) (Top) Schematic of the C_LTM construct used. It is composed of the domain of the antibody λ light chain (C_L) containing a glycosylation site (G1) connected by flexible linkers to the sequence of analysis followed by a C-terminal glycosylation site (G2) and a *c*-myc tag. (Bottom) Scheme depicting the main features of the C_LTM insertion assay. Black dots represent glycosylated sites while white dots represent non-glycosylated sites. (B) Representative halorhodopsin (3QBG) helix G western blot (*n* = 3) of HEK-293T cells transfected with the indicated constructs, and cell lysates were immunoprecipitated with anti-c-myc antiserum in the presence (+) or absence (−) of Endoglycosidase H (EndoH), a glycan-removing enzyme. **C** Representative Ca^{2+} ATPase (1SU4) helix A western blot (*n* = 3) of HEK-293T cells transfected with the indicated constructs, and cell lysates were immunoprecipitated with anti-c-myc antiserum in the presence (+) or absence (-) of EndoH. Non-glycosylated proteins are indicated by two white dots, singly glycosylated proteins are indicated by one white and one black dot, and doubly-glycosylated proteins are indicated by two black dots. Experimental ΔG values (kcal/mol) are shown above each sample (*n* = 3).

sequences containing non-permissive (i, $i + 5$) salt-bridge forming pairs were more efficiently doubly glycosylated (Figures 6(B), lane 3 and 6(C), lane 4, respectively). Thus, similar to what was observed using the Lep system (Figures 4(F) and 5(D)), the presence of paired charges at permissive salt bridge formation distances strongly promotes helix integration into the ER membrane.

Discussion

Charged residues found in α-helices can both provide a stabilizing effect[25] and be important for function such as in zinc finger motifs.[26] Charged residues have been found to be more common in globular helices compared to TM helices,[27] as expected from the hydrophobic environment of the lipid bilayer. In this unfavorable environment both structural changes in the protein (such as helix kinking) or in the membrane (such as membrane bending) may stabilize TM segments despite their high charge content.[28,29] Accordingly, previous experimental data and MD simulations have suggested that when more than one charged residue are found in a TM helix either sequence context or membrane perturbations created by water defects and/or side-chain orientations would lower the energetic penalty of insertion for these heavily charged helices.[30–32]

Pairs of charged residues facing the inside pores in TM regions fulfill an essential functional role, in addition pairs of oppositely charged residues can form salt bridges that can allow them to stably accommodate TM helices when facing the hydrophobic environment.[33] Salt bridges might also be important for shielding the charges during translocon-mediated TM helix insertion.[34] Previous work has shown that charged and polar residues are conserved within TM segments,[18] indicating they are important for function.

In TM-helices Asp-Lys pairs at position i, $i + 4$ and Glu-Lys pairs at position i, $i + 3$ are most prevalent, see Figure 2. In stark contrast to Glu-Arg pair at position i, $i + 1$ they are frequently forming salt-bridges, see Figure 3(B). We noted that approximately 80% of all paired charges found in TM regions are buried within the protein interior and not in direct contact with the membrane core.

An interesting observation is that positively charged Arg-Arg pair at position i, $i + 3$ is among the most common paired charges at this position. Further, positively charged pairs at i, $i + 6$ are also more prevalent than expected and 80% of the Arg-Arg pairs at position i, $i + 6$ also contain an arginine at $i + 3$, as found in helices from voltage-gated ion channels that contain three or more periodically aligned Arg residues with two intervening hydrophobic residues.[35] Interestingly, in our dataset about half of the i, $i + 3$ Arg-Arg pairs are from voltage-gated domains.

The fraction of salt bridges at positions i, $i + 5$ (Figure 3(B)) is an anomaly due to bend α-helices, as found in the helix E from bacterial translocon (PDB ID: 5MG3) and in the helix A from a lipid flippase (PDB ID: 6CC4) due in both cases to the presence of a glycine residue (Figure S6).

By analyzing two native helices containing intra-helical salt bridges we now find that the free energy of insertion (ΔG_{app}) is significantly reduced if both oppositely charged residues are spaced at a permissive distance. These results indicate that intra-helix salt bridge can form during translocon-assisted insertion or even earlier, since in contrast to globular (soluble) helices, TM helices can be compacted inside the ribosome exit tunnel.[36] The reduction in ΔG_{app} seen with Asp-Lys and Asp-Arg pairs in both hR and Ca^{2+} ATPase helices is 0.5–0.7 kcal/mol, which is in good agreement with the 0–1 kcal/mol estimated for these two pairs from thermodynamic peptide partition into octanol experiments.[37] As found in the case of hR helix G (Figures 4(F) and 6B), this reduction might be even higher in the cell context, since some auxiliary components of the membrane insertion machinery[13,38,39] can be not fully represented in the microsomal vesicles preparations.

In hR helix G the lysine residue involved in the salt bridge (K258) plays a fundamental role for protein function (retinal binding) but also introduces a penalty for membrane insertion, a previous event to retinal binding. Our data suggest that the translocon-mediated insertion efficiency is increased by a salt bridge forming between K258 and D254. In the case of Ca^{2+} ATPase helix A the arginine residue involved int the salt bridge (R63) was previously identified from a high throughput computational screen as a residue of interest based on its location in the membrane,[40] thus highlighting how physics-based approaches can lend insights into membrane protein stability.

Beyond the conceptual issues involving the membrane insertion process, we note that the availability of quantitative experimental data on the contribution of salt bridge interactions to the free energy of insertion (ΔG_{app}) will make it possible to fine-tune current membrane protein topology-prediction methods. Topology prediction algorithms based on hidden Markov Models, TMHMM,[41] or hydrophobicity, SCAMPI[42,43] does in general not take salt bridges into account, but methods based on artificial neural networks, MEMSAT,[44,45] could in principle learn this information. Today's state of the art topology prediction tools like the biologically derived scale takes amphiphilicity into account but not salt bridge interactions.[12,46] Current algorithms might tend to overestimate the free energy of insertion due to the penalty imposed by charged residues in the TM region. Distinguishing between paired charges of the same or opposite polarity, i.e., incorporating

the effect of potential salt bridges in the reduction of ΔG_{app} during membrane integration could help to improve prediction tools.

Materials and Methods

Enzymes and chemicals

TNT T7 Quick for PCR DNA was from Promega (Madison, WI, USA). Dog pancreas ER column washed rough microsomes were from tRNA Probes (College Station, TX, USA). EasyTag™ EXPRESS[35]S Protein Labeling Mix, [^{35}S]-L-methionine and [^{35}S]-L-cysteine, for *in vitro* labeling was purchased from Perkin Elmer (Waltham, MA, USA). Restriction enzymes were from New England Biolabs (Massachusetts, USA) and endoglycosidase H was from Roche Molecular Biochemicals (Basel, Switzerland). PCR and plasmid purification kits were from Thermo Fisher Scientific (Ulm, Germany). All oligonucleotides were purchased from Macrogen (Seoul, South Korea).

DNA manipulation

The sequences of interest were introduced into the modified Lep sequence from the pGEM1 plasmid[12] between the *Spe*I and *Kpn*I sites using two double-stranded oligonucleotides with overlapping overhangs at the ends. The complementary oligonucleotides pairs were first annealed at 85 °C for 10 min followed by gradual cooling to 30 °C and ligated into the vector (a kind gift from G. von Heijne's lab). Mutations were obtained by site-directed mutagenesis using the QuikChange kit (Stratagene, La Jolla, California). Lep system including the sequences of interest in the P2 region were subcloned into *Kpn*I linearized pCAGGS in-house version using In-Fusion HD cloning Kit (Takara) according to the manufacturer's instructions. An engineered glycosylation site (Q36N) was added to the C_L-TM plasmid, in which the sequences from hR helix G and ATPase helix A were introduced flanked by 'insulating' Gly-Pro tetrapeptides. A *c*-myc tag (EQKLISEEDL) at the C terminus of the Lep- and C_L-derived sequences was added by PCR before cloning. For *in vitro* assays, DNA was amplified by PCR adding the T7 promoter during the process. All sequences were confirmed by sequencing the plasmid DNA at Macrogen Company (Seoul, South Korea).

Translocon-mediated insertion into microsomal membranes

Lep constructs in pGEM with L4/A15, L5/A14, 3QBG and 1SU4 segments and its variations were transcribed and translated using the TNT T7 Quick Coupled System (#L1170, Promega). Each reaction containing 1 μL of PCR product, 0.5 of EasyTag™ EXPRESS 35S Protein Labeling Mix (Perkin Elmer) (5.5 μCi) and 0.3 μL of microsomes (tRNA Probes) was incubated at 30 °C for 90 min. Endo H treatment was done following the manufacturer's instructions. Samples were analysed by SDS-PAGE (12–14% polyacrylamide). The bands were quantified using a Fuji FLA-3000 phosphoimager and the Image Reader 8.1j software. Free energy was calculated using: $\Delta G_{app} = -RT \ln K_{app}$, where $K_{app} = f_{2g}/f_{1g}$ being f_{1g} and f_{2g} the fraction of singly glycosylated and double glycosylated protein, respectively.

Free apparent insertion energy, ΔG

The free insertion energy of a TM region, ΔG, is calculated as per the experimentally defined Biological hydrophobicity scale.[12] The total ΔG of a region is the sum of individual position specific scores. To note is that hydrophobicity alone is not the only driving force and that the positive inside rule[47,48] and help from preceding TM regions[49,50] can assist insertion in polytopic TM proteins especially.

Core segment definition

We define core segments as a TM region minus the first and last 5 residues. This is to ignore the interface regions which are known to contain polar residues.

Salt bridge definition

Salt bridges are defined as per,[51] where a salt bridge is defined if a side chain carbonyl oxygen atom in Asp-Glu is within 4.0 Å from the nitrogen atom in Arg-Lys. This conforms to other works[7,52] with the definition that the atoms are within hydrogen bond distance. We also define local salt bridges as being bridges that are separated by at most 7 residues in the sequence. This is to separate long salt bridge interactions, which can occur between spatially close residues that are separated in sequence, such as coiled coils where salt bridges can be between separate α-helices.

Transmembrane helices dataset (TM dataset)

The full pipeline is available as a Makefile together with supporting scripts in the GitHub repository. The full PDBTM database[15] was downloaded together with their list of non-redundant protein PDB ids. This list is used to generate both sequence and topology of the proteins by extracting both from the PDBTM-xml. For each protein in the non-redundant list, the membrane regions are extracted using the PDBTM database annotation. All non-membrane regions are annotated 'i' for convenience. To support future analysis, membrane regions longer than 10 were run through DSSP and annotated with 'M' if all residues in

the core segment were defined as α-helix ('H'), otherwise, the full membrane region is annotated 'm'. These proteins are then cluster using cd-hit[53,54] at 40% identity using the parameter -c 0.4 -n 2 -T 0 -M 0 -d 0.

During the extraction of TM regions, the corresponding structure file from RCSB was used to calculate all salt bridges within the current protein and any salt bridge that has at least one residue within any TM region was saved. Additionally, all salt bridges whose both residues were within the same segment and within 7 residues of each other were annotated as local as per the salt bridge definition above.

Extraction of charged residues

From all annotated TM regions of length 17 or longer,[2] the core segment was extracted. All these core regions were then scanned and when a charged residue was encountered, we recorded any other charged residue from 7 residues before the current one to 7 residues after. This results in charged residues that can contain a charged pairing partner outside of the TM segment and will therefore differ slightly from charged pairs which are defined next.

Extraction of charged pairs

From all annotated TM regions of length 17 or longer,[2] the core segment was extracted. All these core regions were then scanned and when a charged residue was encountered, we look at up to 7 residues in front of it or to the end of that core region, whichever came first. All occurrences of charged pairs were recorded, resulting in charged pairs where both residues were fully within the core segment of a TM helix.

MSA-dataset extension (TM-MSA)

Using the TM dataset, we extended it by creating an MSA alignment of each protein using jackhmmer[55] against Uniref90 with one iteration and an E-value cut off of 10^{-3} with the following parameters: -N 1 -E 1e-3 --incE 1e-3 --cpu 14.

From each alignment, we then sampled up to 200 hits, including the initial seed sequence. If there were fewer than 200 hits, we used them all. We then used the original topology for each alignment to extract all TM regions, only to include parts where the sequence covers the full TM region and where the sequence did not contain any insertions or deletions.

Dataset of helices from globular proteins (GLOB and GLOB-MSA)

To create a reference dataset of globular α-helical proteins we extracted all globular all-α-protein domains from SCOP. As SCOP classifies domains of proteins resulting in that one domain

of a protein can be annotated as globular whereas another domain as TM (see 1PPJ chain D as an example) we reduced the SCOP list against the redundant list of all PDBTM[15] chains to clear any overlap. This results in 4500 proteins in total.

Topological files with sequence and membrane topology are created with the help of the RCSB secondary structure file and only membrane segments whose core region (*i.e.*, central 15 residues) is annotated as pure (canonical) α-helices were retained, *i.e.*, those TM segments containing any residues within the core annotated as loops or other types of secondary structures were removed. This file was further homology reduced and alignments prepared in the same manner and using the same parameters as the TM dataset described above.

The GLOB-MSA dataset was created in the same way as the TM-MSA dataset using jackhmmer to extend the sequences to alignments and then to extract helix sequences.

Redundant (TM-Red) dataset

The full PDBTM database was used as in the preparation of the TM dataset. We skip the clustering step and instead use all redundant proteins to generate their respective topology files. We added in the constraint that each selected TM region must fully contain at least one potential salt bridge. This means a local salt bridge where both residues are within the core segment. This dataset was only used to find potential candidates for further *in vitro* and *in whole cell* experiments. See section of natural salt bridges above.

Calculation of log odds ratio

The log odds ratios for each amino acid pair for steps 1 through 10 are calculated as follows (using the natural logarithm):

$$logOddsRatio = log((A/B)/(C/D))$$

where for two amino acids p_1 and p_2:

A = number of pairs of p_1 to p_2
B = number of total pairs
C = number of p_1 times number of p_2
D = number of total pairs squared
A *logOddsRatio* of 2, or -2, corresponds to a $7.3\times$ increase/decrease, respectively.

The standard error, SE, and z-value is calculated as follows allowing for a two-sided test:

$$SE = \sqrt{(1/A + 1/B + 1/C + 1/D)}$$

$$z = abs(logOddsValue/SE)$$

The survival function, sf, from the scipy python packages is used to calculate the *p*-values. To

correct for multiple hypothesis, the Bonferroni Correction is used based on the number of hypotheses, 20 * 20 * 10, number of amino acids square times the number of steps.

Expression in mammalian cells

Lep or C_L-derived constructs containing 3QBG or 1SU4 segments and its variations were tagged with *c*-myc epitope at their Ct (EQKLISEEDL) and inserted in the appropriate plasmids. Once the sequence was verified, plasmids were transfected into HEK293-T cells using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Approximately 24 h post-transfection cells were harvested and washed with PBS buffer. After a short centrifugation (1000 rpm for 5 min on a table-top centrifuge) cells were lysed by adding 100 μL of lysis buffer (30 mM Tris–HCl, 150 mM NaCl, 0.5% Nonidet P-40) were sonicated in an ice bath in a bioruptor (Diagenode) during 10 min and centrifuged. After protein quantification, equal amounts of protein were submitted to Endo H treatment or mock-treated followed by SDS-PAGE analysis and transferred into a PVDF transfer membrane (ThermoFisher Scientific) as previously described.[56] Protein glycosylation status was analysed by Western Blot using an anti-*c*-myc antibody (Sigma), anti-rabbit IgG-peroxidase conjugated (Sigma), and with ECL developing reagent (GE Healthcare). Chemiluminescence was visualized using an ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare).

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

† Equal contribution

References

1. Martínez-Gil, L., Saurí, A., Marti-Renom, M.A., Mingarro, I., (2011). Membrane protein integration into the endoplasmic reticulum. *FEBS J.* **278**, 3846–3858. https://doi.org/10.1111/j.1742-4658.2011.08185.x.

2. Baeza-Delgado, C., Marti-Renom, M.A., Mingarro, I., (2012). Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* **42**, 199–207. https://doi.org/10.1007/s00249-012-0813-9.

3. Tsirigos, K.D., Peters, C., Shu, N., Käll, L., Elofsson, A., (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407. https://doi.org/10.1093/nar/gkv485.

4. Tsirigos, K.D., Govindarajan, S., Bassot, C., Västermark, Å., Lamb, J., Shu, N., et al., (2018). ScienceDirect Topology of membrane proteins — predictions, limitations and variations. *Curr. Opin. Struct. Biol.* **50**, 9–17. https://doi.org/10.1016/j.sbi.2017.10.003.

5. Bañó-Polo, M., Baeza-Delgado, C., Orzáez, M., Marti-Renom, M.A., Abad, C., Mingarro, I., (2012). Polar/Ionizable Residues in Transmembrane Segments: Effects on Helix-Helix Packing. *PLoS ONE* **7**, https://doi.org/10.1371/journal.pone.0044263.t001 e44263.

6. Marqusee, S., Baldwin, R.L., (1987). Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898–8902.

7. Donald, J.E., Kulp, D.W., DeGrado, W.F., (2011). Salt bridges: Geometrically specific, designable interactions. *Proteins* **79**, 898–915. https://doi.org/10.1002/prot.22927.

8. Mbaye, M.N., Hou, Q., Basu, S., Teheux, F., Pucci, F., Rooman, M., (2019). A comprehensive computational study of amino acid interactions in membrane proteins. *Sci. Rep.* **9**, 12043–12114. https://doi.org/10.1038/s41598-019-48541-2.

9. Baeza-Delgado, C., von Heijne, G., Marti-Renom, M.A., Mingarro, I., (2016). Biological insertion ofcomputationally designed shorttransmembrane segments. *Sci. Rep.*, 1–9. https://doi.org/10.1038/srep23397.

10. Chin, C.N., von Heijne, G., (2000). Charge pair interactions in a model transmembrane helix in the ER membrane. *J.*

Mol. Biol. **303**, 1–5. https://doi.org/10.1006/jmbi.2000.4122.

11. Braunger, K., Pfeffer, S., Shrimal, S., Gilmore, R., Berninghausen, O., Mandon, E.C., et al., (2018). Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215–219. https://doi.org/10.1126/science.aar7899.

12. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., et al., (2005). Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381. https://doi.org/10.1038/nature03216.

13. Tamborero, S., Vilar, M., Martínez-Gil, L., Johnson, A.E., Mingarro, I., (2011). Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J. Mol. Biol.* **406**, 571–582. https://doi.org/10.1016/j.jmb.2011.01.009.

14. Bañó-Polo, M., Martínez-Gil, L., Barrera, F.N., Mingarro, I., (2019). Insertion of Bacteriorhodopsin Helix C Variants into Biological Membranes. *ACS Omega* **5**, 556–560. https://doi.org/10.1021/acsomega.9b03126.

15. Kozma, D., Simon, I., Tusnády, G.E., (2013). PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524–D529.

16. Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., Murzin, A.G., (2013). SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res.* **42**, D310–D314. https://doi.org/10.1093/nar/gkt1242.

17. Andreeva, A., Kulesha, E., Gough, J., Murzin, A.G., (2019). The SCOP database in 2020: expanded classification of representative family and superfamily domains of known protein structures. *Nucleic Acids Res.* **48**, D376–D382. https://doi.org/10.1093/nar/gkz1064.

18. Illergård, K., Kauko, A., Elofsson, A., (2011). Why are polar residues within the membrane core evolutionary conserved? *Proteins* **79**, 79–91. https://doi.org/10.1002/prot.22859.

19. Samanta, U., Pal, D., Chakrabarti, P., (1999). Packing of aromatic rings against tryptophan residues in proteins. *Acta Cryst.* **D55**, 1421–1427. https://doi.org/10.1107/S090744499900726X.

20. Kanada, S., Takeguchi, Y., Murakami, M., Ihara, K., Kouyama, T., (2011). Crystal Structures of an O-Like Blue Form and an Anion-Free Yellow Form of pharaonis Halorhodopsin. *J. Mol. Biol.* **413**, 162–176.

21. Kumar, S., Nussinov, R., (2002). Relationship between Ion Pair Geometries and Electrostatic Strengths in Proteins. *Biophys. J.* **83**, 1595–1612. https://doi.org/10.1016/S0006-3495(02)73929-5.

22. Toyoshima, C., Nakasako, M., Nomura, H., Ogawa, H., (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature* **405**, 647–655. https://doi.org/10.1038/35015017.

23. Kauko, A., Hedin, L.E., Thebaud, E., Cristobal, S., Elofsson, A., von Heijne, G., (2010). Repositioning of transmembrane alpha-helices during membrane protein folding. *J. Mol. Biol.* **397**, 190–201. https://doi.org/10.1016/j.jmb.2010.01.042.

24. Feige, M.J., Hendershot, L.M., (2013). Quality Control of Integral Membrane Proteins by Assembly-Dependent Membrane Integration. *Mol. Cell.* **51**, 297–309. https://doi.org/10.1016/j.molcel.2013.07.013.

25. Armstrong, K.M., Baldwin, R.L., (1993). Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 11337–11340. https://doi.org/10.1073/pnas.90.23.11337.

26. Lin, C.-Y., Lin, L.-Y., (2018). The conserved basic residues and the charged amino acid residues at the α-helix of the zinc finger motif regulate the nuclear transport activity of triple C2H2 zinc finger proteins. *PLoS ONE* **13**, e0191971–e192020. https://doi.org/10.1371/journal.pone.0191971.

27. Kauko, A., Illergård, K., Elofsson, A., (2008). Coils in the Membrane Core Are Conserved and Functionally Important. *J. Mol. Biol.* **380**, 170–180. https://doi.org/10.1016/j.jmb.2008.04.052.

28. Callenberg, K.M., Latorraca, N.R., Grabe, M., (2012). Membrane bending is critical for the stability of voltage sensor segments in the membrane. *J. Gen. Physiol.* **140**, 55–68. https://doi.org/10.1085/jgp.201110766.

29. Guerriero, C.J., Gomez, Y.K., Daskivich, G.J., Reutter, K.-R., Augustine, A.A., Weiberth, K.F., et al., (2019). Harmonizing Experimental Data with Modeling to Predict Membrane Protein Insertion in Yeast. *Biophys. J.* **117**, 668–678. https://doi.org/10.1016/j.bpj.2019.07.013.

30. Dorairaj, S., Allen, T.W., (2007). On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4943–4948. https://doi.org/10.1073/pnas.0610470104.

31. Martínez-Gil, L., Pérez-Gil, J., Mingarro, I., (2008). The Surfactant Peptide KL4 Sequence Is Inserted with a Transmembrane Orientation into the Endoplasmic Reticulum Membrane. *Biophys. J.* **95**, L36–L38. https://doi.org/10.1529/biophysj.108.138602.

32. MacCallum, J.L., Bennett, W.F.D., Tieleman, D.P., (2011). Transfer of arginine into lipid bilayers is nonadditive. *Biophys. J.* **101**, 110–117. https://doi.org/10.1016/j.bpj.2011.05.038.

33. Walther, T.H., Ulrich, A.S., (2014). Transmembrane helix assembly and the role of salt bridges. *Curr. Opin. Struct. Biol.* **27**, 63–68. https://doi.org/10.1016/j.sbi.2014.05.003.

34. Whitley, P., Grau, B., Gumbart, J.C., Martínez-Gil, L., Mingarro, I., (2021). Folding and Insertion of Transmembrane Helices at the ER. *IJMS* **22**, 12778. https://doi.org/10.3390/ijms222312778.

35. Okamura, Y., Fujiwara, Y., Sakata, S., (2015). Gating Mechanisms of Voltage-Gated Proton Channels. *Annu. Rev. Biochem.* **84**, 685–709. https://doi.org/10.1146/annurev-biochem-060614-034307.

36. Bañó-Polo, M., Baeza-Delgado, C., Tamborero, S., Hazel, A., Grau, B., Nilsson, I., et al., (2018). Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nature Commun.* **9**, 5246. https://doi.org/10.1038/s41467-018-07554-7.

37. Jayasinghe, S., Hristova, K., White, S.H., (2001). Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **312**, 927–934. https://doi.org/10.1006/jmbi.2001.5008.

38. Shurtleff, M.J., Itzhak, D.N., Hussmann, J.A., Schirle Oakdale, N.T., Costa, E.A., Jonikas, M., et al., (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife* **7**, 382. https://doi.org/10.7554/eLife.37018.

39. Chitwood, P.J., Hegde, R.S., (2020). An intramembrane chaperone complex facilitates membrane protein biogenesis. *Nature* **584**, 1–22. https://doi.org/10.1038/s41586-020-2624-y.

40. Marcoline, F.V., Bethel, N., Guerriero, C.J., Brodsky, J.L., Grabe, M., (2015). Membrane Protein Properties Revealed through Data-Rich Electrostatics Calculations. *Structure* **23**, 1526–1537. https://doi.org/10.1016/j.str.2015.05.014.

41. Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L. L., (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. https://doi.org/10.1006/jmbi.2000.4315.

42. Viklund, H., Elofsson, A., (2008). OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**, 1662–1668. https://doi.org/10.1093/bioinformatics/btn221.

43. Peters, C., Tsirigos, K.D., Shu, N., Elofsson, A., (2016). Improved topology prediction using the terminal hydrophobic helices rule. *Bioinformatics* **32**, 1158–1162. https://doi.org/10.1093/bioinformatics/btv709.

44. Nugent, T., Jones, D.T., (2009). Transmembrane protein topology prediction using support vector machines. *BMC Bioinf.* **10**, 980–1011. https://doi.org/10.1186/1471-2105-10-159.

45. Nugent, T., Jones, D.T., (2010). Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. *PLoS Comput. Biol.* **6**, https://doi.org/10.1371/journal.pcbi.1000714 e1000714.

46. Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., et al., (2007). Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030. https://doi.org/10.1038/nature06387.

47. von Heijne, G., (1989). Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456–458. https://doi.org/10.1038/341456a0.

48. Lerch-Bader, M., Lundin, C., Kim, H., Nilsson, I., von Heijne, G., (2008). Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *PNAS* **105**, 4127–4132. https://doi.org/10.1073/pnas.0711580105.

49. Hedin, L.E., Ojemalm, K., Bernsel, A., Hennerdal, A., Illergård, K., Enquist, K., et al., (2010). Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. *J. Mol. Biol.* **396**, 221–229. https://doi.org/10.1016/j.jmb.2009.11.036.

50. Bañó-Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J.L., Elofsson, A., Mingarro, I., (2013). Charge Pair Interactions in Transmembrane Helices and Turn Propensity of the Connecting Sequence Promote Helical Hairpin Insertion. *J. Mol. Biol.* **425**, 830–840. https://doi.org/10.1016/j.jmb.2012.12.001.

51. Kumar, S., Ma, B., Tsai, C.J., Nussinov, R., (2000). Electrostatic strengths of salt bridges in thermophilic and mesophilic glutamate dehydrogenase monomers. *Proteins* **38**, 368–383. https://doi.org/10.1002/(sici)1097-0134(20000301)38:4<368::aid-prot3>3.0.co;2-r.

52. Bosshard, H.R., Marti, D.N., Jelesarov, I., (2004). Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **17**, 1–16. https://doi.org/10.1002/jmr.657.

53. Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152.

54. Li, W., Godzik, A., (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22** (13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158.

55. Eddy, S.R., (2011). Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **7**, e1002195

56. Duart, G., García-Murria, M.J., Grau, B., Acosta-Cáceres, J.M., Martínez-Gil, L., Mingarro, I., (2020). SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol.* **10**, 200209–200216. https://doi.org/10.1098/rsob.200209.

57. Engelman, D.M., Steitz, T.A., Goldman, A., (1986). Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321–353. https://doi.org/10.1146/annurev.bb.15.060186.001541.

58. Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., Lomize, A.L., (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376. https://doi.org/10.1093/nar/gkr703.

A

AA	Gap	Sequence	$\Delta G_{pred.}$	$\Delta G_{exp.}$
L5/A14	-	AAAALALAALAALALAAAA	-1.00	-1.17 ± 0.12
K	-	AAAALALKALAALALAAAA	1.00	0.31 ± 0.08
D	-	AAAALALAALADLALAAAA	0.98	0.57 ± 0.08
KD	3	AAAALALAKLADLALAAAA	2.97	0.69 ± 0.03
KD	4	AAAALALKALADLALAAAA	2.93	0.58 ± 0.05
KD	5	AAAALAKLALADLALAAAA	2.14	0.88 ± 0.04

B
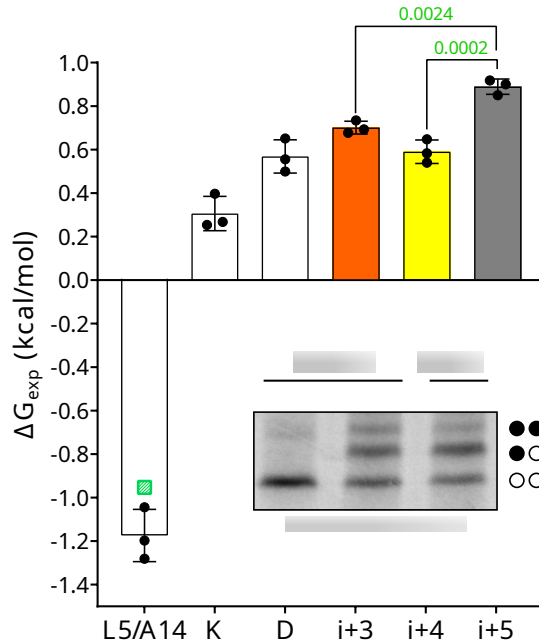
Figure S1. Effects on membrane insertion of single or pairs of Asp and Lys residues in L5/A14. A The tested sequences from L5/A14 model TM (including the charged residues, bold), the gap distance, and the predicted ΔG (ΔG_{pred}) and experimental (ΔG_{exp}) values in kcal/mol are shown. Amino acids with positive and negative charge are highlighted in blue (K) and red (D) respectively. B Experimental ΔG (ΔG_{exp}) in kcal/mol of each tested sequence in the Lep-based microsomal assay. The mean and standard deviation of 3 independent experiments are represented. The individual value of each experiment is represented by a solid dot, p-values are indicated above. In addition, a green dot represents the ΔG_{pred} value for the L5/A14 sequence. The wt and single mutants are shown in white bars. Charges at compatible distances with salt bridge formation (i, i+3; and i, i+4) are shown in orange and yellow, respectively. Not compatible distances with salt bridge formation (i, i+5) is shown in gray. The inset shows a representative SDS-PAGE gel for L4/A15 and L5/A14 constructs. The construct was expressed in rabbit reticulocyte lysed in the presence (+RM) or absence (−RM) of column washed rough microsomes. Bands of non-glycosylated proteins are indicated by a white dot; mono and double glycosylated proteins are indicated by one and two black dots, respectively.

Figure S2
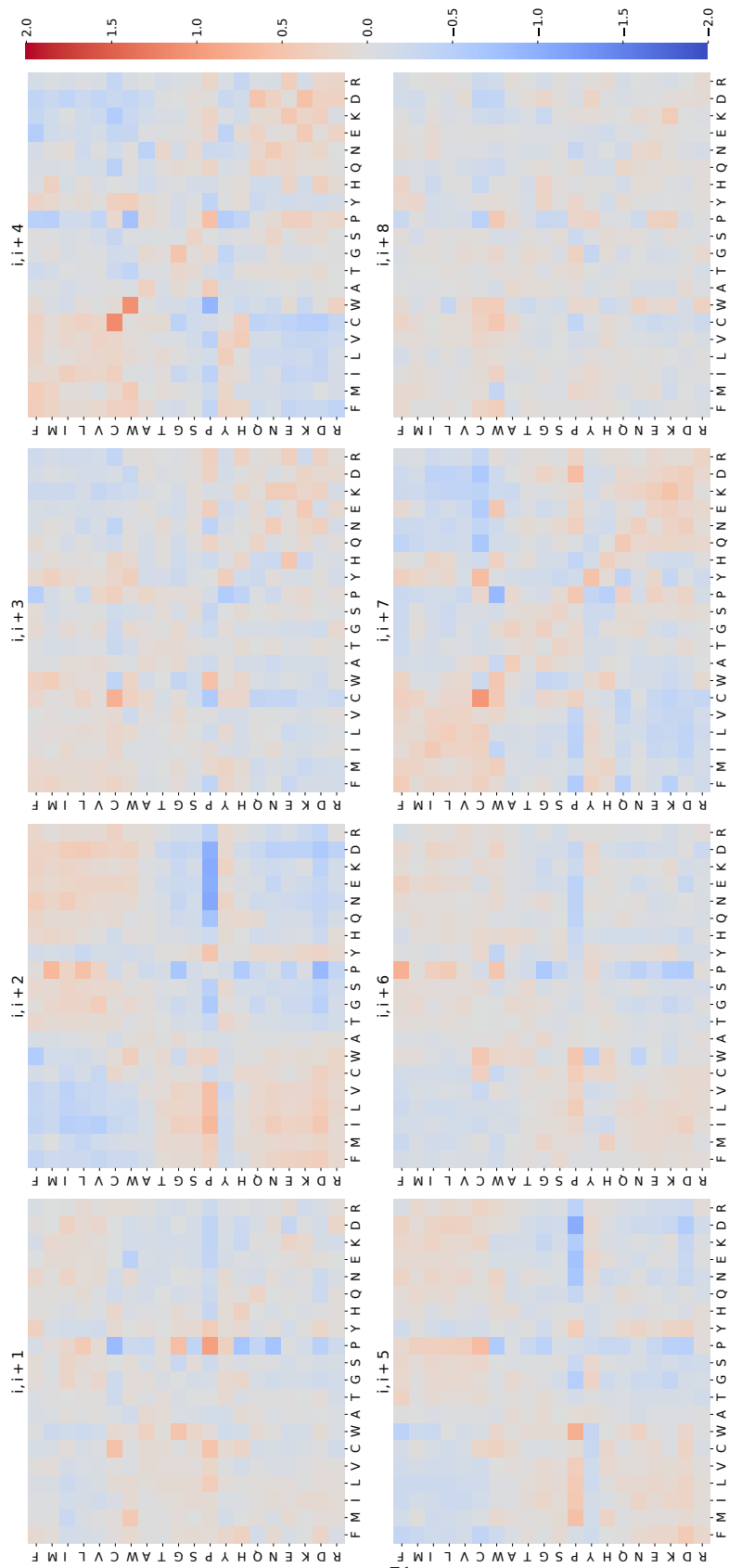
Figure S2. Log odds ratios of each pair of amino acids for 'i, i+1' through 'i, i+8' for the GLOB-MSA dataset. Log odds ratios for the middle core residues of α-helices of at least 17 residues in length in the GLOB-MSA dataset. The rows on the y-axis indicate the first amino acid in the pair and the columns on the x-axis the second. The residues are ordered as in Fig. 2.
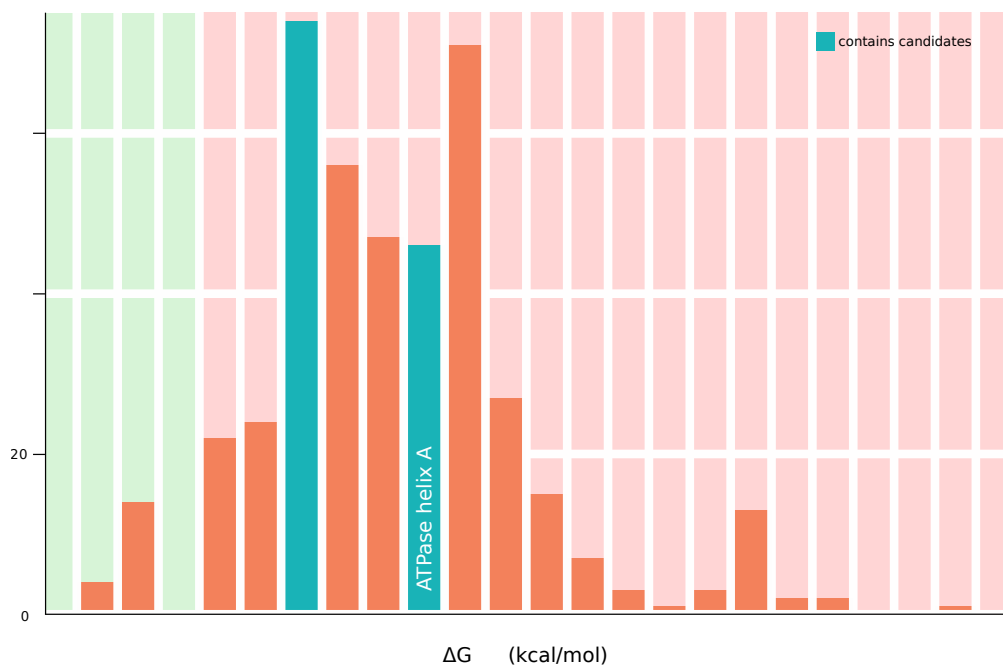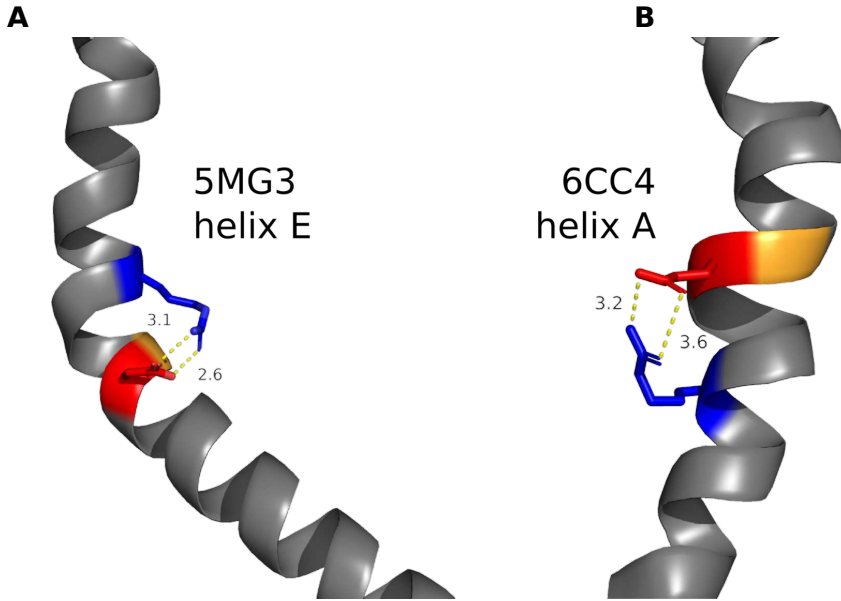
Figure S3
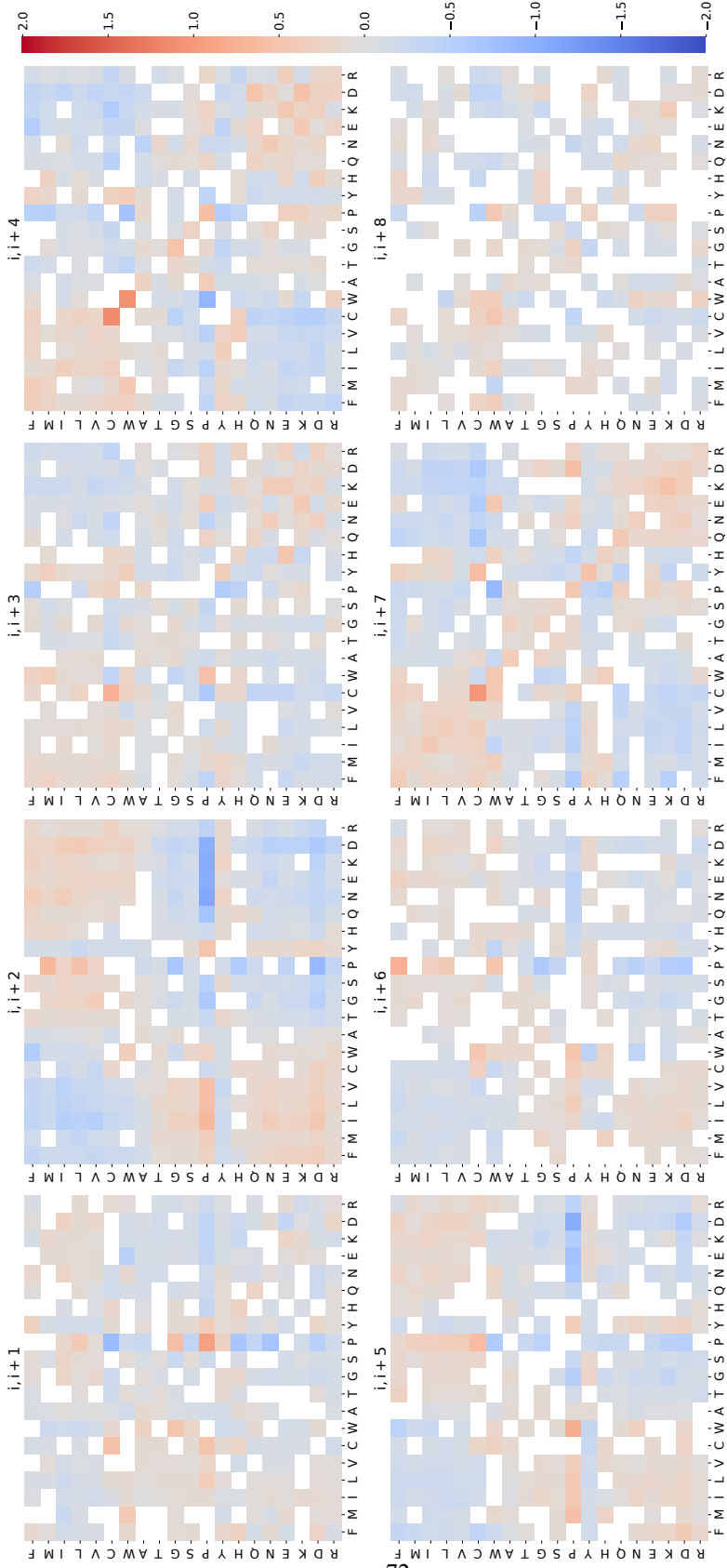
Figure S3. Log odds ratios of each pair of amino acids for 'i, i+1' through 'i, i+8' for the GLOB-MSA dataset. Log odds ratios as in S2 but all pairs with P-value >0.05 have been masked.

72

Figure S4

Figure S4. Masked log odds ratios of each pair of amino acids for 'i, i+1' through 'i, i+8' for the TM-MSA dataset. Log odds ratios as in Fig. 2 but all pairs with P-value >0.05 have been masked.

Figure S5

Figure S5. Histogram of ΔG$_{pred}$ values for the TM segments containing paired charges. ΔG values represented with a bin size of 0.8 kcal/mol, with negative (indicative of insertion) and positive (indicative of non-inserted) values highlighted with a green and red background, respectively. Halorhodopsin 3QBG helix G (ΔGpred of 1.73 kcal/mol) and Ca^{2+} ATPase helix A (ΔG$_{pred}$ of 4.15 kcal/mol) are part of the left and right blue highlighted bars, respectively.

Figure S6. Salt bridges at pair i, i+5. A helix E from bacterial translocon (PDB ID: 5MG3). Glycine (orange) residue causes a kink facilitating salt bridge interaction between Asp50 (red) and Arg55 (blue). **B** Helix A from a lipid flippase (PDB ID: 6CC4). Glycine (orange) residue causes a kink facilitating salt bridge interaction between Arg153 (blue) and Glu158 (red). Distances are shown in Ångström.

4.2. Chapter 2: Topology

SARS-CoV-2 envelope protein topology in eukaryotic membranes

Gerard Duart[†], Mª Jesús García-Murria[†], Brayan Grau[†], 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
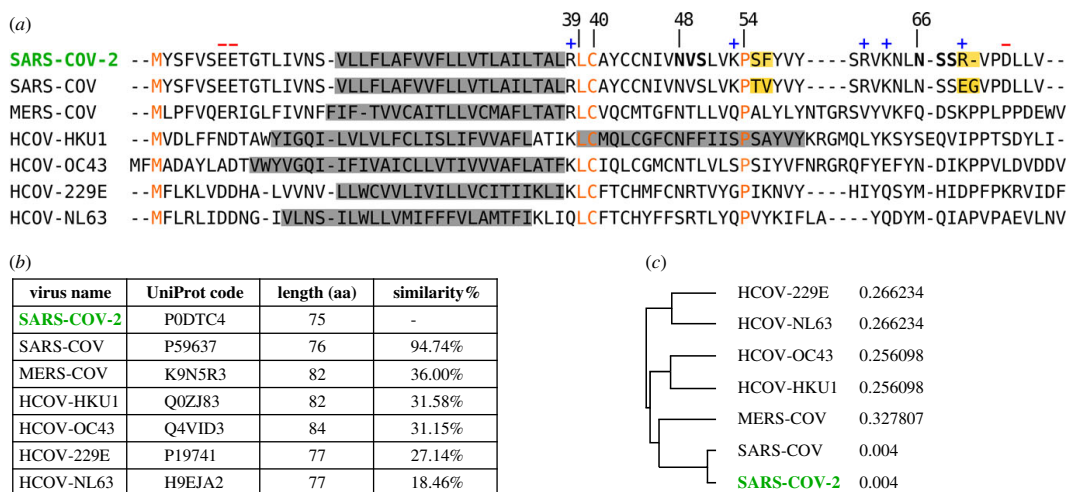
MJG-M, 0000-0002-8396-0335; BG, 0000-0002-8582-023X; LM-G, 0000-0002-9076-7760; IM, 0000-0002-1910-1229

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 (Nt_{lum}/Ct_{cyt}). 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 hydropathicity (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 1*b*), followed by MERS-CoV (36.00%). Interestingly, sequence similarities are significantly lower for the other four human coronaviruses, which usually cause mild to moderate

(a)

```
                                                    39 40      48      54              66
                                                    + |        |       |               +
                                                    -- -                               + + +      + -
SARS-COV-2   --MYSFVSEETGTLIVNS-VLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPSFYVY----SRVKNLN-SSR-VPDLLV--
SARS-COV     --MYSFVSEETGTLIVNS-VLLFLAFVVFLLVTLAILTALRLCAYCCNIVNVSLVKPTVYVY----SRVKNLN-SSEGVPDLLV--
MERS-COV     --MLPFVQERIGLFIVNFFIF-TVVCAITLLVCMAFLTATRLCVQCMTGFNTLLVQPALYLYNTGRSVYVKFQ-DSKPPLPPDEWV
HCOV-HKU1    --MVDLFFNDTAWYIGQI-LVLVLFCLISLIFVVAFLATIKIMQLCGFCNFFIISPSAYVYKRGMQLYKSYSEQVIPPTSDYLI-
HCOV-OC43    MFMADAYLADTVWYVGQI-IFIVAICLLVTIVVVAFLATFKLCIQLCGMCNTLVLSPSIYVFNRGRQFYEFYN-DIKPPVLDVDDV
HCOV-229E    --MFLKLVDDHA-LVVNV-LLWCVVLIVILLVCITIIKLIKLCFTCHMFCNRTVYGPIKNVY----HIYQSYM-HIDPFPKRVIDF
HCOV-NL63    --MFLRLIDDNG-IVLNS-ILWLLVMIFFFVLAMTFIKLIQLCFTCHYFFSRTLYQPVYKIFLA----YQDYM-QIAPVPAEVLNV
```

(b)

virus name	UniProt code	length (aa)	similarity %
SARS-COV-2	P0DTC4	75	-
SARS-COV	P59637	76	94.74%
MERS-COV	K9N5R3	82	36.00%
HCOV-HKU1	Q0ZJ83	82	31.58%
HCOV-OC43	Q4VID3	84	31.15%
HCOV-229E	P19741	77	27.14%
HCOV-NL63	H9EJA2	77	18.46%

(c)

HCOV-229E	0.266234
HCOV-NL63	0.266234
HCOV-OC43	0.256098
HCOV-HKU1	0.256098
MERS-COV	0.327807
SARS-COV	0.004
SARS-COV-2	0.004

Figure 1. (*a*) Multi-alignment of amino acid sequences of the E protein of SARS-CoV-2 and the other six human coronavirus. SARS-CoV severe acute respiratory syndrome coronavirus (UniProt P59637), MERS-CoV Middle East respiratory syndrome coronavirus (UniProt K9N5R3), HCoV-HKU1 (UniProt Q0ZJ83), HCoC-OC43 (UniProt Q4VID3), HCoC-229E (UniProt P19741) and HCoV-NL63 (UniProt Q5SBN7). Predicted TM segments at UniProt are highlighted in a grey box. Native predicted glycosylation acceptor sites in SARS-CoV-2 are shown in bold and charged residues highlighted with + or − symbols on top. Conserved residues are shown in orange. Differences between SARS-CoV-2 and SARS-CoV are highlighted as yellow boxes. (*b*) Phylogenetic data and (*c*) tree obtained with Clustal Omega (EMBL-EBI) using the default parameters.

upper-respiratory tract illness typical for common cold, with the lowest similarity found for HCOV-NL63 (18.46%). These findings are consistent with the phylogenetic tree proposed based on the amino acid sequences of the human coronavirus E proteins using ClustalW (figure 1*c*).

2. Results and discussion

2.1. E protein topology prediction

Computer-assisted analysis of the SARS-CoV-2 E protein amino acid sequence using seven popular prediction methods showed that all membrane protein prediction algorithms except MEMSAT-SVM suggested the presence of one transmembrane (TM) segment located roughly around amino acids 12 to 39 (table 1), which is not predicted as a cleavable signal sequence according to SignalP-5.0 [7]. Regarding E protein topology, TMHMM and Phobius predicted an N-terminus cytosolic orientation, while MEMSAT-SVM, TMpred, HMMTop and TOPCONS predicted an N-terminus luminal orientation. These discrepancies found among the predictions from different algorithms motivated experimental approaches.
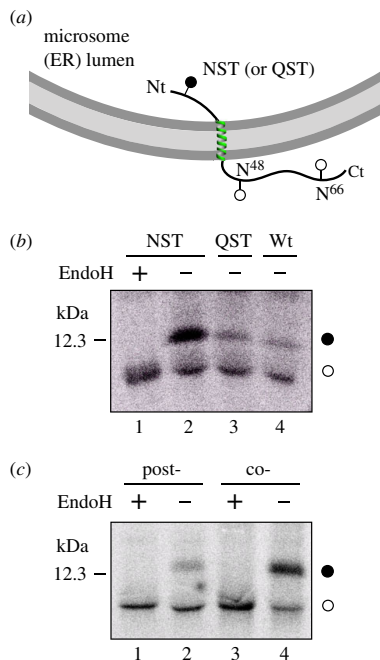
2.2. Insertion into microsomal membranes

First, we performed *in vitro* E protein transcription/translation experiments in the presence of ER-derived microsomes and [^{35}S]-labelled amino acids. The membrane insertion orientation of the predicted TM segment into microsomal membranes was based on N-linked glycosylation and summarized in figure 2*a*. N-linked glycosylation has been extensively used as topological reporter for more than two decades [8]. In eukaryotic cells, proteins can only be glycosylated in the lumen of the ER because the active site of oligosaccharyl transferase (OST), a translocon-associated protein responsible for N-glycosylation

[9], is located there [10]; no N-linked glycosylation occurs within the membrane or in the cytosol. It is important to note that two possible N-linked glycosylation sites are located C-terminally of the predicted TM segment in E protein wild-type sequence at positions N48 and N66 (figure 1). However, N48 is not expected to be modified even if situated lumenally due to the close proximity of this glycosylation acceptor site to the membrane if the hydrophobic region is recognized as TM by the translocon [11,12]. Thus, mono-glycosylation (at N66) would serve as a C-terminal translocation reporter. To test N-terminal translocation, a construct was engineered where a predicted highly efficient glycosylation acceptor site (i.e. Asn-Ser-Thr, NST) was designed at the N-terminus. When E protein constructs were translated *in vitro* in the presence of microsomes, the protein was significantly glycosylated when the N-terminal designed glycosylation site was present, as shown by the increase in the electrophoretic mobility of the slower radioactive band after an endoglycosidase H (Endo H) treatment (figure 2*b*, lanes 1 and 2). However, when a control (Gln-Ser-Thr, QST) that is not a glycosylation acceptor site (lane 3) or the wild-type (lane 4) sequences were translated, E protein molecules were minimally glycosylated. Since multiple topologies have been reported for previous coronavirus E proteins [13–17], SARS-CoV-2 E protein insertion into the microsomal membranes in two opposite orientations cannot be discounted, but according to our data being dominant an Nt$_{lum}$/Ct$_{cyt}$ orientation.

2.3. E protein integrates cotranslationally into microsomal membranes

We have previously reported that several viral membrane proteins are cotranslationally inserted into ER-derived microsomal membranes [18–20]. Since membrane protein insertion and N-glycosylation are coupled at the ER by complex

Figure 2. Translocon-mediated insertion of E protein variants into microsomal membranes. (*a*) Schematic representation of E protein constructs. Glycosylation acceptor Asn residues are indicated. (*b*) *In vitro* translation in the presence of microsomes of the different E protein constructs. Construct containing inserted asparagine and threonine residues at positions 3 and 5 (NST; lanes 1–2) or glutamine and threonine at positions 3 and 5 (lane 3), and wild-type variants (lane 4) were translated in the presence of microsomes. NST variant was split and half of the sample was Endo H treated (lane 1). Bands of non-glycosylated and glycosylated proteins are indicated by white and black dots, respectively. (*c*) E protein (harbouring an engineered glycosylation site at the N-terminus, positions 3–5) was translated in either the absence (lanes 1 and 2) or the presence (lanes 3 and 4) of microsomal membranes. In lanes 1 and 2, microsomal membranes were added posttranslationally (after 1 h, post-) and incubation was continued for another 1 h. Samples in lanes 1 and 3 were treated later with EndoH. The gels are representative of at least three independent experiments.

Table 1. Computer analysis of the SARS-CoV-2 E protein amino acid sequence topology. n.p., non-predicted.

algorithm	Nt	Ct	TMDs (start-end)
ΔG predictor	n.p.	n.p.	1 (17–39)
TMHMM	cytosol	lumen	1 (12–34)
MEMSAT-SVM	lumen	lumen	2 (10–39) (43–58)
TMpred	lumen	cytosol	1 (17–34)
HMMTop	lumen	cytosol	1 (11–35)
Phobius	cytosol	lumen	1 (12–37)
TOPCONS	lumen	cytosol	1 (16–36)

formation of a ribosome, the translocon and the OST [10], we sought to investigate whether or not SARS-CoV-2 E protein is cotranslationally inserted into the ER membrane by blocking protein synthesis after E protein (harbouring N-terminal NST
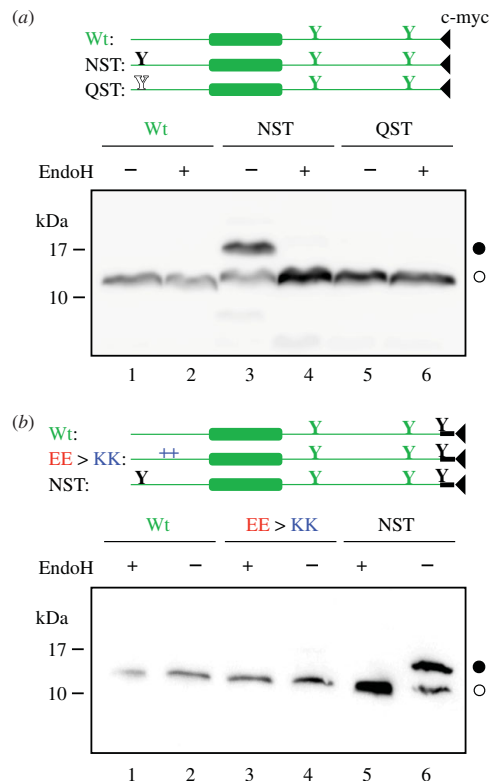
Figure 3. E protein topology in mammalian cells. To determine the topology *in vivo* HEK-293T cells were transfected with C-terminal tagged (c-myc) E protein variants. (*a*) Constructs encoding wild-type (Wt; lanes 1 and 2), inserted asparagine and threonine at positions 3 and 5 (NST; lanes 3 and 4) or glutamine and threonine at positions 3 and 5 (QST; lanes 5 and 6) were Endo H (+) or mock (−) treated. Filled and empty Y-shaped symbols denoted acceptor (NST) and non-acceptor (QST) glycosylation sites, respectively. (*b*) Additionally, we included constructs containing similar Wt (lanes 1 and 2), replaced glutamic acids at positions 7 and 8 by lysine residues (EE > KK; lanes 3 and 4) or NST (lanes 5 and 6) variants with an extra glycosylation site inserted at the Ct end of the protein. Once again, to confirm the glycosylated nature of the higher molecular weight bands, samples were either Endo H (+) or mock (−) treated. Designed glycosylation sites and tags are shown in black, while native E protein features are shown in grey.

glycosylation site) has been translated in the absence of membranes. As shown in figure 2*c*, E protein (NST) was efficiently glycosylated when microsomal membranes were added to the translation mixture cotranslationally (lane 4). But when microsomal membranes were included posttranslationally after the translation was inhibited by cycloheximide, the protein was only residually glycosylated (lane 2), suggesting that E protein is mainly integrated cotranslationally through the ER translocon. This means that the microsomal insertion machinery recognizes, orients and provides a path into the membrane for this viral protein.

2.4. Membrane topology in mammalian cells

To analyse protein topology in mammalian cells, a series of E protein variants tagged with c-myc epitope at the C-terminus

were transfected into HEK-293T cells. As shown in figure 3a, only an E protein construct harbouring the N-terminal engineered acceptor site was efficiently modified (lanes 1–4), denoting an N-terminal ER luminal localization (Nt_{lum}). Several topological parameters have been proposed to govern membrane protein topology, among which the preferential distribution of positively charged residues in the cytosol (positive-inside rule) has been established as the primary topology determinant both experimentally [21] and statistically [22]. E protein is a single-spanning membrane protein with an even net charge distribution on both sides of the membrane. There are only eight charged residues along the protein sequence (two negatively charged residues preceding the TM segment, and five positively and one negatively charged residues at the C-terminal domain; figure 1a), which correlates well the observed topology with the 'positive-inside rule'. However, negatively charged residues have also been proved to significantly affect the topology [23]. To test the robustness of the observed topology, we added an optimized Ct glycosylation tag [24] and replaced the two negatively charged residues located in the translocated N-terminal domain (E7 and E8) by two lysine residues (figure 3b). In cells expressing this mutant E protein (EE > KK), the protein retained its C-terminal tail at the cytosolic side of the membrane as indicated by the absence of glycosylated forms (figure 3b, lanes 3 and 4). These data reveal that topological determinants have only a minor effect on viral membrane protein topology as previously demonstrated for other viruses [25] and suggest that viral membrane protein topology could have co-evolved with the protein environment of its natural host, ensuring proper membrane protein orientation. Altogether, the present *in vivo* results demonstrate that SARS-CoV-2 E protein is a single-spanning membrane protein with an Nt_{lum}/Ct_{cyt} orientation in mammalian cell membranes. Similarly, SARS-CoV E protein was shown to mainly adopt an Nt_{lum}/Ct_{cyt} topology in the infected cell and mammalian cells expressing the recombinant protein [26]. This topology is compatible with the ion channel capacity described previously [27], and with the recently published pentameric structural model of SARS-CoV E protein in micelles [28], in which the C-terminal tail of the protein is α-helical and extramembrane.

3. Concluding remarks

The membrane topology described here would allow the cytoplasmic C-terminal tail of the E protein to interact with the C-termini of M and/or S SARS-CoV-2 membrane-embedded proteins [3], and/or with Golgi scaffold proteins as previously described for other coronaviruses [29], to induce virus budding or influence vesicular traffic through the Golgi complex by collecting viral membrane proteins for assembly at Golgi membranes. Future experiments will have to unravel whether these functions involve the SARS-CoV-2 E protein.

4. Material and methods

4.1. Enzymes and chemicals

TNT T7 Quick for PCR DNA was from Promega (Madison, WI, USA). Dog pancreas ER rough microsomes were from tRNA Probes (College Station, TX, USA). EasyTag EXPRESS35S

Protein Labeling Mix, [35S]-L-methionine and [35S]-L-cysteine, for *in vitro* labelling was purchased from Perkin Elmer (Waltham, MA, USA). Restriction enzymes were from New England Biolabs (Massachusetts, USA) and endoglycosidase H was from Roche Molecular Biochemicals (Basel, Switzerland). PCR and plasmid purification kits were from Thermo Fisher Scientific (Ulm, Germany). All oligonucleotides were purchased from Macrogen (Seoul, South Korea).

4.2. Computer-assisted analysis of E protein sequence

Prediction of transmembrane segments was done using up to 7 of the most common methods available on the Internet: ΔG Predictor [30,31] (http://dgpred.cbr.su.se/), TMHMM [32] (http://www.cbs.dtu.dk/services/TMHMM/), MEMSAT-SVM [33] (http://bioinf.cs.ucl.ac.uk/psipred/), TMpred (https://embnet.vital-it.ch/software/TMPRED_form.html), HMMTop [34] (http://www.enzim.hu/hmmtop/), Phobius [35] (http://phobius.sbc.su.se/) and TOPCONS [36] (http://topcons.net/). All user-adjustable parameters were left at their default values.

4.3. DNA manipulation

Full-length E protein was synthesized by Invitrogen (*GeneArt* gene synthesis) and subcloned into *Kpn*I linearized pCAGGS in-house version [37] using In-Fusion HD cloning Kit (Takara) according to the manufacturer's instructions. For *in vitro* assays, DNA was amplified by PCR adding the T7 promoter and the relevant glycosylation sites during the process. N-terminal NST glycosylation site was designed by inserting an asparagine and a threonine before and after Ser3, respectively. Control no-glycosylable QST site was introduced in similarly inserting a glutamine residue instead of an asparagine. All E protein variants were obtained by site-directed mutagenesis using QuikChange kit (Stratagene, La Jolla, California) and were confirmed by sequencing the plasmid DNA at Macrogen Company (Seoul, South Korea).

4.4. Translocon-mediated insertion into microsomal membranes

E protein variants, PCR amplified from pCAGGS, were transcribed and translated using the TNT T7 Quick for PCR DNA coupled transcription/translation system (Promega, USA). The reactions contained 10 µl of TNT, 2 µl of PCR product, 1 µl of EasyTag (5 µC$_i$) and 0.6 µl of column-washed microsomes (tRNA Probes, USA) and were incubated for 60 min at 30°C. Translation products were ultracentrifuged ($100\,000g$ for 15 min) on a 0.5 M sucrose cushion and analysed by SDS-PAGE. For the endoglycosidase H (Endo H), the treatment was done as previously described [20]. Briefly, the translation mixture was diluted in 120 µl of PBS and centrifuged on a 0.5 M sucrose cushion ($100\,000g$ 15 min 4°C). The pellet was then suspended in 50 µl of sodium citrate buffer with 0.5% SDS and 1% β-mercaptoethanol, boiled 5 min, and incubated 1 h at 37°C with 1 unit of Endo H. Then, the samples were analysed by SDS-PAGE and gels were visualized on a Fuji FLA3000 phosphorimager using Image Reader 8.1j software.

4.5. Postranslational and cotranslational insertion assay

E protein DNAs were transcribed/translated (30°C 1 h) either in the absence (figure 2*c*, post- samples) or in the presence (co- samples) of microsomal membranes. The translation was inhibited with cycloheximide (10 min, 26°C, 2 mg ml^{-1} final concentration), after which microsomes were added to those samples labelled as posttranslational and incubated for an additional hour at 30°C. Subsequently, membranes were collected by ultracentrifugation; half of the samples were EndoH treated and analysed by SDS-PAGE (double volume was loaded for the post-samples due to the lower translation levels observed). Protein molecules were visualized on a Fuji FLA3000 phosphorimager.

4.6. E protein expression in mammalian cells

E protein sequence variants were tagged with a c-myc epitope at their C-terminus (Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu, EQKLISEEDL) and inserted in a pCAGGS-ampicillin plasmid. When appropriate (figure 3*b*), an optimized glycosylation site followed by a flexible dipeptide (Asn-Ser-Thr-Gly-Ser, NSTGS) [24,38] preceded the c-myc epitope. Once the sequence was verified, plasmids were transfected into HEK293-T cells using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Approximately 24 h post-transfection cells were harvested and washed with PBS buffer. After short centrifugation (1000 r.p.m. for 5 min on a table-top centrifuge), cells were lysed by adding 100 µl of lysis buffer (30 mM Tris-HCl, 150 mM NaCl, 0.5% Nonidet P-40), sonicated in an ice bath in a bioruptor (Diagenode) during 10 min and were centrifugated. Total protein was quantified and equal amounts of protein submitted to Endo H treatment or mock-treated, followed by SDS-PAGE analysis and transferred into a PVDF transfer membrane (ThermoFisher Scientific). Protein glycosylation status was analysed by Western Blot using an anti-c-myc antibody (Sigma), anti-rabbit IgG-peroxidase conjugated (Sigma) and with ECL developing reagent (GE Healthcare). Chemiluminescence was visualized using an ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare).

References

1. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. 2020 Embracing the landscape of therapeutics. *Cell* **181**, 1–19. (doi:10.1016/j.cell.2020.03.025)

2. Wu A *et al.* 2020 Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe* **27**, 325–328. (doi:10.1016/j.chom.2020.02.001)

3. Schoeman D, Fielding BC. 2019 Coronavirus envelope protein: current knowledge. *Virol. J.* **16**, 1–22. (doi:10.1186/s12985-019-1182-0)

4. Ruch TR, Machamer CE. 2012 The coronavirus E protein: assembly and beyond. *Viruses* **4**, 363–382. (doi:10.3390/v4030363)

5. Almazán F, DeDiego ML, Sola I, Zuñiga S, Nieto-Torres JL, Marquez-Jurado S, Andrés G, Enjuanes L. 2013 Engineering a replication-competent, propagation-defective middle east respiratory syndrome coronavirus as a vaccine candidate. *mBio* **4**, 20282-11. (doi:10.1128/mBio.00650-13)

6. Netland J, DeDiego ML, Zhao J, Fett C, Álvarez E, Nieto-Torres JL, Enjuanes L, Perlman S. 2010 Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. *Virology* **399**, 120–128. (doi:10.1016/j.virol.2010.01.004)

7. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Heijne G, Nielsen H. 2019 The blind babymaker. *Nat. Biotechnol.* **37**, 1–8. (doi:10.1038/nbt.4341)

8. Mingarro I, von Heijne G, Whitley P. 1997 Membrane-protein engineering. *Trends Biotechnol.* **15**, 432–437. (doi:10.1016/S0167-7799(97)01101-3)

9. Martínez-Gil L, Saurí A, Marti-Renom MA, Mingarro I. 2011 Membrane protein integration into the endoplasmic reticulum. *FEBS J* **278**, 3846–3858. (doi:10.1111/j.1742-4658.2011.08185.x)

10. Braunger K, Pfeffer S, Shrimal S, Gilmore R, Berninghausen O, Mandon EC, Becker T, Förster F, Beckmann R. 2018 Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215–219. (doi:10.1126/science.aar7899)

11. Nilsson IM, von Heijne G. 1993 Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem.* **268**, 5798–5801.

12. Orzáez M, Salgado J, Giménez-Giner A, Pérez-Payá E, Mingarro I. 2004 Influence of proline residues in transmembrane helix packing. *J. Mol. Biol.* **335**, 631–640. (doi:10.1016/j.jmb.2003.10.062)

13. Corse E, Machamer CE. 2000 Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J. Virol.* **74**, 4319–4326. (doi:10.1128/JVI.74.9.4319-4326.2000)

14. Raamsman MJ, Locker JK, de Hooge A, de Vries AA, Griffiths G, Vennema H, Rottier PJ. 2000 Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *J. Virol.* **74**, 2333–2342. (doi:10.1128/JVI.74.5.2333-2342.2000)

15. Godet M, L'Haridon R, Vautherot JF, Laude H. 1992 TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* **188**, 666–675. (doi:10.1016/0042-6822(92)90521-P)

16. Maeda J, Repass JF, Maeda A, Makino S. 2001 Membrane topology of coronavirus E protein. *Virology* **281**, 163–169. (doi:10.1006/viro.2001.0818)

17. Yuan Q, Liao Y, Torres J, Tam JP, Liu DX. 2006 Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. *FEBS Lett* **580**, 3192–3200. (doi:10.1016/j.febslet.2006.04.076)

18. Saurí A, Saksena S, Salgado J, Johnson AE, Mingarro I. 2005 Double-spanning plant viral movement protein integration into the endoplasmic reticulum membrane is signal recognition particle-dependent, translocon-mediated, and concerted. *J. Biol. Chem.* **280**, 25 907–25 912. (doi:10.1074/jbc.M412476200)

19. Martínez-Gil L, Saurí A, Vilar M, Pallás V, Mingarro I. 2007 Membrane insertion and topology of the p7B movement protein of melon necrotic spot virus

(MNSV). *Virology* **367**, 348–357. (doi:10.1016/j.virol.2007.06.006)

20. Martínez-Gil L, Johnson AE, Mingarro I. 2010 Membrane insertion and biogenesis of the *Turnip Crinkle Virus* p9 movement protein *J. Virol*. **11**, 5520–5527. (doi:10.1128/JVI.00125-10)

21. von Heijne G. 1989 Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456–458. (doi:10.1038/341456a0)

22. Baeza-Delgado C, Marti-Renom MA, Mingarro I. 2013 Structure-based statistical analysis of transmembrane helices. *Eur. Biophys. J.* **42**, 199–207. (doi:10.1007/s00249-012-0813-9)

23. Nilsson I, von Heijne G. 1990 Fine-tuning the topology of a polytopic membrane protein: role of positively and negatively charged amino acids. *Cell* **62**, 1135–1141. (doi:10.1016/0092-8674(90)90390-Z)

24. Bañó-Polo M, Baldin F, Tamborero S, Marti-Renom MA, Mingarro I. 2011 *N*-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* **20**, 179–186. (doi:10.1002/pro.551)

25. Saurí A, Tamborero S, Martínez-Gil L, Johnson AE, Mingarro I. 2009 Viral membrane protein topology is dictated by multiple determinants in its sequence. *J. Mol. Biol.* **387**, 113–128. (doi:10.1016/j.jmb.2009.01.063)

26. Nieto-Torres JL, DeDiego ML, Álvarez E, Jimenez-Guardeño JM, Regla-Nava JA, Llorente M, Kremer L, Shuo S, Enjuanes L. 2011 Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein. *Virology* **415**, 69–82. (doi:10.1016/j.virol.2011.03.029)

27. Verdiá-Báguena C, Nieto-Torres JL, Alcaraz A, DeDiego ML, Torres J, Aguilella VM, Enjuanes L. 2012 Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. *Virology* **432**, 485–494. (doi:10.1016/j.virol.2012.07.005)

28. Surya W, Li Y, Torres J. 2018 Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1860**, 1309–1317. (doi:10.1016/j.bbamem.2018.02.017)

29. Corse E, Machamer CE. 2003 The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology* **312**, 25–34. (doi:10.1016/S0042-6822(03)00175-2)

30. Hessa T, Meindl-Beinker NM, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White SH, von Heijne G. 2007 Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030. (doi:10.1038/nature06387)

31. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. 2005 Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381. (doi:10.1038/nature03216)

32. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. 2001 Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580. (doi:10.1006/jmbi.2000.4315)

33. Nugent T, Jones DT. 2009 Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 980-11. (doi:10.1186/1471-2105-10-159)

34. Tusnády GE, Simon I. 2001 The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850. (doi:10.1093/bioinformatics/17.9.849)

35. Käll L, Krogh A, Sonnhammer ELL. 2004 A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036. (doi:10.1016/j.jmb.2004.03.016)

36. Tsirigos KD, Peters C, Shu N, Käll L, Elofsson A. 2015 The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407. (doi:10.1093/nar/gkv485)

37. Vera-Velasco NM, García-Murria MJ, del Pino MMS, Mingarro I, Martínez-Gil L. 2018 Proteomic composition of Nipah virus-like particles. *J. Proteomics* **172**, 190–200. (doi:10.1016/j.jprot.2017.10.012)

38. Tamborero S, Vilar M, Martínez-Gil L, Johnson AE, Mingarro I. 2011 Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J. Mol. Biol.* **406**, 571–582. (doi:10.1016/j.jmb.2011.01.009)

BBA Research Letter

The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness

Gerard Duart, Maria J. García-Murria, Ismael Mingarro [*]

Departament de Bioquímica i Biologia Molecular, Institut Universitari de Biotecnologia i Biomedicina (BIOTECMED), Universitat de València, E-46100 Burjassot, Spain

ABSTRACT

- Single-spanning SARS-CoV-2 envelope (E) protein topology is a major determinant of protein quaternary structure and function.
- Charged residues distribution in E protein sequences from highly pathogenic human coronaviruses (*i.e.*, SARS-CoV, MERS-CoV and SARS-CoV-2) stabilize Nt_{out}-Ct_{in} membrane topology.
- E protein sequence could have evolved to ensure a more robust membrane topology from MERS-CoV to SARS-CoV and SARS-CoV-2.

In the past 20 years, the world has seen three human coronaviruses responsible for severe disease outbreaks: the Severe Acute Respiratory Syndrome (SARS-CoV) that emerged in 2002, the Middle East Respiratory Syndrome (MERS-CoV) in 2012 and recently the emergence of SARS-CoV-2, which has spread around the world at an unprecedented rate, causing a worldwide pandemic.

Coronaviruses' genome includes four major structural proteins: membrane (M), spike (S), nucleocapsid (N) and envelope (E). The multifunctional E protein is the smallest of the structural proteins (between 8 and 12 kDa) and has the lowest copy number in the lipid envelope of mature virus particles [1]. The majority of the E protein pool localizes to the endoplasmic reticulum Golgi intermediate compartment (ERGIC) in the host cell where it participates in virus budding, assembly and trafficking [2]. In addition to this structural role the E protein oligomerizes to form pentameric ion channels similar to viroporins [3–5] and possesses a C-terminal PDZ-binding motif that induce immunopathology by overexpression of inflammatory cytokines [6]. These features of E protein play a major role in the exacerbated immune response causing the acute respiratory syndrome, the leading cause of death in SARS-CoV and SARS-CoV-2 [7], and have been shown to be critical for propagation of other human coronaviruses. The assembly of E protein into the ER membrane in the correct orientation (topology) is critical for its functions. In the evolution of membrane proteins it is not rare to observe mutations leading to a more fixed orientation relative to the membrane.

SARS-CoV-2 E protein is a single-spanning membrane protein with a skewed distribution of charged residues on both sides of the membrane. There are only eight charged residues in the protein sequence, two negatively charged residues N-terminal to the transmembrane (TM) domain, and five positively plus one negatively charged residues in the C-terminal domain (Fig. 1A). The observed Nt_{out}-Ct_{in} topology [8] is in good agreement with the 'positive-inside' rule [9].

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 [8]. Interestingly, sequence similarities are significantly higher for the coronaviruses that usually cause severe illness than for those that cause mild to moderate upper-respiratory tract symptoms typical for common cold. SARS-CoV-2 E protein has the highest similarity to SARS-CoV (94.74%) with only minor differences (Fig. 1A), followed by MERS-CoV (36.00%) [8]. Nevertheless, regarding topology determination, there is a common feature in all of them. There is a positively charged residue strongly conserved and strategically located proximal to the hydrophobic region C-terminal end (Supplementary Fig. 1) in all seven human coronaviruses. It is worth mentioning that this positively charged residue is an Arg (Arg38) in MERS-CoV, SARS-CoV and SARS-CoV-2 (Fig. 1A), while in the other human coronaviruses is a Lys [8]. Interestingly, in the analysis of 81,818 E protein sequences of SARS-CoV-2 globally available, no change was detected in

this position [10]. Positively charged residues located near the cytoplasmic end of hydrophobic segments in membrane proteins promote correct membrane insertion of TM helices. It has been determined that a single Arg or Lys residue typically contributes approximately −0.5 kcal/mol to the apparent free energy of membrane insertion when placed at this location [11].

In comparison to globular (water-soluble) proteins, topology provides an extra dimension that membrane proteins can evolve. Topology can evolve, for example, by redistribution of charged residues on both sides of the membrane. The alignment of MERS-CoV, SARS-CoV and SARS-CoV-2 E proteins unveils a tendency to accumulate a net positive charge balance C-terminally to the TM domain (Fig. 1A), which
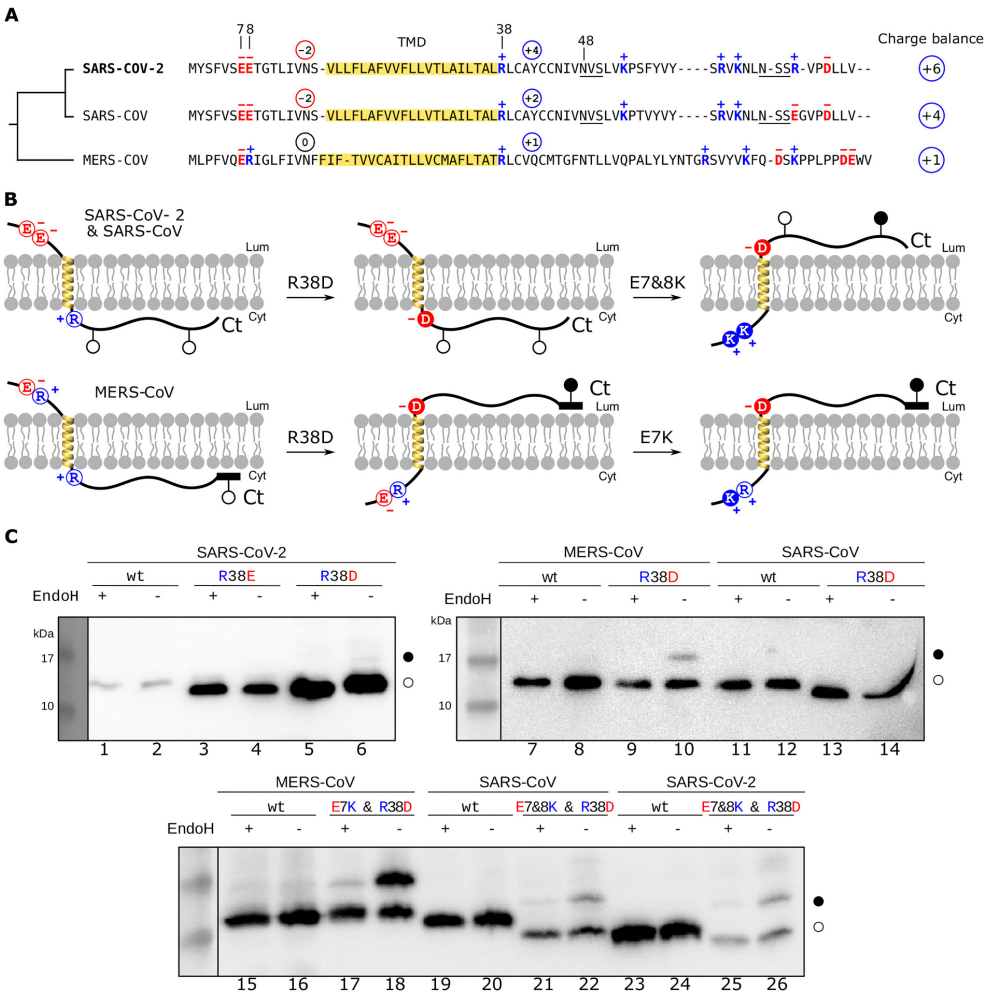
Fig. 1. A. Multi-alignment of amino acid sequences of the E protein from MERS-CoV (UniProt K9N5R3), SARS-CoV (UniProt P59637) and SARS-CoV-2 (UniProt P0DTC4). Predicted TM segments are highlighted in a yellow box. Negatively charged amino acids are shown in red with – symbol on top while the positive ones are shown in blue with + symbol on top. Native predicted glycosylation acceptor sites are underlined. Conserved and relevant residues are marked with the number on top (7, 8, 38 and 48). The net charge summation before and after the TM segment is shown encircled. The charge balance (charge balance at the region following the TM segment minus charge balance at the region preceding the TM segment) is shown at the right side. Tree obtained with Clustal Omega (EMBL-EBI) using the default parameters.

B. Schematic representations of E protein topology in the presence of the different mutations. Wild type residues 7, 8 and 38 are shown in an empty colored circle (red for glutamic acids and blue for arginines) accompanied with − or + symbol depending on the charge of the residue. Point mutations are shown in red (negative) or blue (positive) solid circles emphasizing the charge change. Glycosylation acceptor sites are indicated with white (non-glycosylated) or black (glycosylated) dots. In MERS-CoV, Ct-tail containing the glycosylation site is represented with a black rectangle.

C. To determine the topology *in vivo*, HEK-293T cells were transfected with Ct tagged (c-myc) E protein variants. The E protein virus and the proper mutations are indicated on top of each gel. Lanes with odd numbers are Endo H treated (+) and even numbers are mock treated (−). Samples were separated on SDS-PAGE (14% polyacrylamide) and analyzed by Western blot using an anti-*c*-myc antibody (Sigma). Bands of non-glycosylated and glycosylated proteins are indicated by white and black dots, respectively. The gels are representative of at least three independent experiments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

correlates with the 'positive-inside' rule, but also suggests an increasing robustness in the topology determination from MERS-CoV to SARS-CoV-2. MERS-CoV E protein sequence contains one positively and one negatively charged residues in the translocated N-terminus and four positively charged residues plus three negatively charged residues in the C-terminal cytosolic domain, giving a net charge balance of +1. In the case of SARS-CoV, charge balance increases substantially with a net charge of −2 in the N-terminal extra-membranous domain and +2 (4 positively plus 2 negatively charged residues) in the C-terminus, giving a net charge balance of +4. In the case of SARS-CoV-2, this balance is higher due to E69R substitution, giving a net charge balance of +6 (Fig. 1A).

The topology of the SARS-CoV-2 envelope protein was recently proved to be Nt_{out}-Ct_{in} in eukaryotic membranes [8]. To test the topological relevance of the conserved Arg38 residue we designed two replacement mutants in which the positively charged residue was mutated to aspartic or glutamic acid residues (R38D or R38E, respectively). The topology was determined by monitoring glycosylation of the consensus acceptor sites that the E protein has downstream of the TM segment (Fig. 1A). Glycosylation at a single site increases the molecular weight of the protein by ~2.5 kDa. In eukaryotic cells, proteins can only be glycosylated in the lumen of the ER because the active site of oligosaccharyl transferase, the enzyme responsible for co-translational glycosylation, is located there. To analyze protein topology in mammalian cells, E protein variants tagged with *c-myc* epitope at the C-terminus were transfected into HEK-293T cells. As shown in Fig. 1C, neither R38E mutant (lanes 3 and 4) nor R38D mutant (lanes 5 and 6) resulted on alteration of the original E protein topology (lanes 1 and 2). The N-terminal translocation of these mutants was demonstrated by engineering two highly efficient glycosylation sites, one at the N-terminus and another one in a C-terminal tag (Supplementary Fig. 2). Similarly, R38D mutation in SARS-CoV E protein displayed the same glycosylation pattern as the wild-type equivalent (Fig. 1C, lanes 11–14). The MERS-CoV E protein sequence does not contain natural glycosylation consensus acceptor sites (Fig. 1A). Therefore an optimized C-terminal glycosylation tag was added to the C-terminal domain (Fig. 1B) [12]. In this case, no glycosylation band was observed when the wild-type protein was expressed (Fig. 1C, lanes 7 and 8). However, a higher molecular weight band was detected when the R38D mutant was expressed (lane 10). The nature of the higher molecular weight protein species was confirmed by endoglycosidase H (EndoH) treatment (lane 9), a highly specific enzyme that cleaves oligosaccharides regardless of their location. Thus, in the case of MERS-CoV E protein some inverted molecules were observed when R38D mutant was expressed. This replacement eliminates the positive charge balance at the C-terminal domain. These data reveal that topological determinants have only a limited effect on viral membrane protein topology as previously observed for other viruses [13] and suggests that E protein in coronaviruses could have evolved to ensure a more robust membrane topology.

Recent statistical studies have suggested that negatively charged residue enrichment in the non-cytoplasmic regions can modulate membrane protein topology [14]. To challenge the robustness of the topology observed for E proteins, we decided to replace the negatively charged residues found in the translocated N-termini in combination with the designed R38D mutations. In the case of MERS-CoV E protein there is only one negatively charged residue (Glu7) in the N-terminal domain, while both SARS-CoV and SARS-CoV-2 E proteins have two (Glu7 and Glu8, Fig. 1A). The combination mutant (E7K & R38D) showed a stronger topology effect on MERS-CoV E protein, since this protein was strongly glycosylated when expressed (Fig. 1C, lanes 15–18). SARS-CoV and SARS-CoV-2 E proteins with the combined mutations (E7&8K & R38D) had only a small proportion of molecules with the reversed topology, suggesting stronger topology determination (Fig. 1C, lanes 19–26), especially if we take into account that the observed effect is generated by a triple mutation. It should be mentioned that the consensus glycosylation acceptor site at Asn48 is not expected to be modified even if situated luminally due to its close proximity to the membrane [8].

In all three cases, the conserved Arg38 residue plays a limited role in the topology determination. However, its relevance is likely ameliorated with other topological determinants in human coronavirus E protein sequences. Our results suggest that viral evolution has played an important role in strengthening the E protein (Nt_{out}-Ct_{in}) topology from MERS-CoV to SARS coronaviruses. Probably, the R8E mutation present in both SARS-CoVs compared with MERS-CoV is one of the factors contributing to topology robustness, by converting the net charge of 0 at N-terminal region of MERS-CoV into a −2 in both SARS-CoVs, in good agreement with the "negative outside enrichment" rule suggested from statistics derived from a large body of membrane protein sequences [14] and observed in membrane protein structures [15]. At the same time, an evolutionary tendency to accumulate positively charged residues in the cytoplasmic C-terminal domain of these E proteins could be observed (Fig. 1A), contributing to a multifactorial effect on membrane topology, which allows quaternary protein structure formation [4,5] and plays an essential role in viral infection and pathogenesis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbamem.2021.183608.

References

[1] Y.M. Bar-On, A. Flamholz, R. Phillips, R. Milo, SARS-CoV-2 (COVID-19) by the numbers, eLife 9 (2020), 1787-15.

[2] J.L. Nieto-Torres, et al., Subcellular location and topology of severe acute respiratory syndrome coronavirus envelope protein, Virology 415 (2011) 69–82.

[3] C. Verdiá-Báguena, et al., Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids, Virology 432 (2012) 485–494.

[4] W. Surya, Y. Li, J. Torres, Structural model of the SARS coronavirus E channel in LMPG micelles, Biochim. Biophys. Acta Biomembr. 1860 (2018) 1309–1317.

[5] V.S. Mandala, et al., Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers, Nat. Struct. Mol. Biol. 27 (2020) 1–24.

[6] J.M. Jimenez-Guardeño, et al., The PDZ-binding motif of severe acute respiratory syndrome coronavirus envelope protein is a determinant of viral pathogenesis, PLoS Pathog. 10 (2014) e1004320.

[7] I. Alam, et al., Functional pangenome analysis shows key features of E protein are preserved in SARS and SARS-CoV-2, Front. Cell. Infect. Microbiol. 10 (2020) e82210–e82219.

[8] G. Duart, et al., SARS-CoV-2 envelope protein topology in eukaryotic membranes, Open Biol. 10 (2020), 200209-6.

[9] G. von Heijne, Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues, Nature 341 (1989) 456–458.

[10] M.S. Rahman, et al., Mutational insights into the envelope protein of SARS-CoV-2, Gene Rep. 22 (2021) 100997.

[11] M. Lerch-Bader, C. Lundin, H. Kim, I. Nilsson, G. von Heijne, Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum 105 (2008) 4127–4132.

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

[13] A. Saurí, S. Tamborero, L. Martínez-Gil, A.E. Johnson, I. Mingarro, Viral membrane protein topology is dictated by multiple determinants in its sequence, J. Mol. Biol. 387 (2009) 113–128.

[14] J.A. Baker, W.-C. Wong, B. Eisenhaber, J. Warwicker, F. Eisenhaber, Charged residues next to transmembrane regions revisited: "positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule", BMC Biol. 15 (2017) 1–29.

[15] C. Baeza-Delgado, M.A. Marti-Renom, I. Mingarro, Structure-based statistical analysis of transmembrane helices, Eur. Biophys. J. 42 (2012) 199–207.

Supplementary information

Materials and Methods

Enzymes and chemicals. Restriction enzymes were from New England Biolabs (Massachusetts, USA) and endoglycosidase H was from Roche Molecular Biochemicals (Basel, Switzerland). PCR and plasmid purification kits were from Thermo Fisher Scientific (Ulm, Germany). All oligonucleotides were purchased from Macrogen (Seoul, South Korea).

DNA Manipulation. Full-length E protein sequences from MERS-CoV, SARS-CoV and SARS-CoV-2 were synthesized by Invitrogen (*GeneArt* gene synthesis) and subcloned into *Kpn*I linearized pCAGGS using In-Fusion HD cloning Kit (Takara) according to the manufacturer's instructions. N-terminal NST glycosylation site in SARS-CoV-2 E protein (Supplementary Fig. 2) was designed by inserting an asparagine and a threonine before and after Ser3, respectively. All E protein variants were obtained by site-directed mutagenesis using QuikChange kit (Stratagene, La Jolla, California) and were confirmed by sequencing the plasmid DNA at Macrogen Company (Seoul, South Korea).

E protein expression in mammalian cells. E protein sequence variants were tagged with an optimized C-terminal glycosylation site [1,2] plus a c-myc epitope at their C-terminus and inserted in a pCAGGS-ampicillin plasmid. Once the sequence was verified, plasmids were transfected into HEK293-T cells using Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. Approximately 24 h post-transfection cells were harvested and washed with PBS buffer. After a short centrifugation (1000 rpm for 5 min on a table-top centrifuge) cells were lysed by adding 100 µL of lysis buffer (30 mM Tris-HCl, 150 mM NaCl, 0.5% Nonidet P-40) were sonicated on ice using a bioruptor (Diagenode) during 10min and centrifugated. Total proteins were quantified and equal amounts of protein submitted to Endoglycosidase H (EndoH) treatment or mock-treated, followed by SDS-PAGE analysis (18% polyacrylamide) and transferred into a PVDF transfer membrane (ThermoFisher Scientific). Protein glycosylation status was analysed by Western Blot using an anti-c-myc antibody (Sigma), anti-rabbit IgG-peroxidase conjugated (Sigma), and with ECL developing reagent (GE Healthcare). Chemiluminescence was visualized using an ImageQuantTM LAS 4000mini Biomolecular Imager (GE Healthcare).
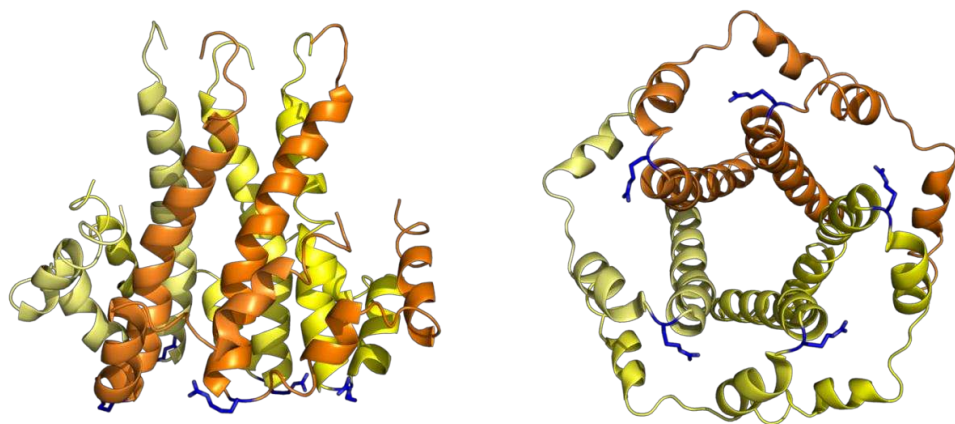
Supp. Fig. 1. Lateral (left) and cytoplasmic (virion-inside) (right) views of the pentameric structure of SARS-CoV E protein (PDB code: 5X29) [3] showing each monomer from light yellow to dark orange with Arg38 side chain as sticks shown in blue.

Supplementary Figure 2
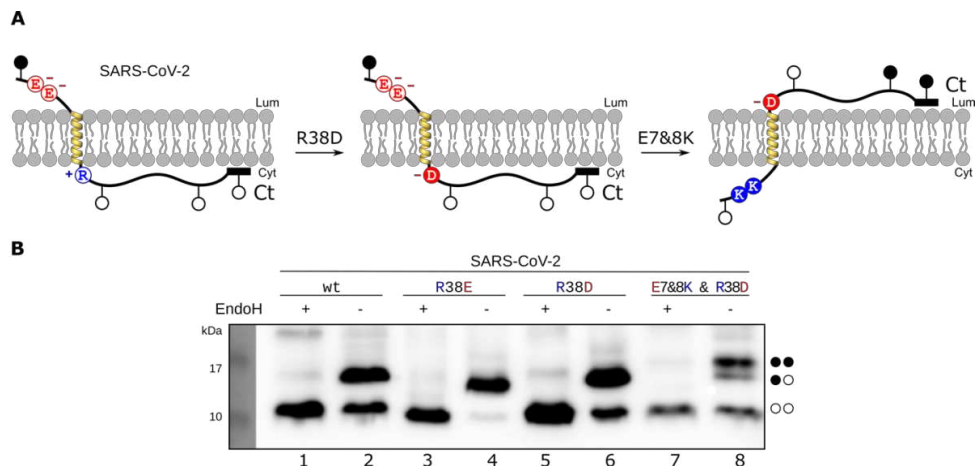
Supp. Fig. 2. A. Schematic representations of SARS-CoV-2 E protein topology in the presence of the different mutations and two engineered glycosylation sites (Nt and Ct) as previously [4]. Wild type residues 7, 8 and 38 are shown in an empty colored circle (red for glutamic acids and blue for arginine) with − or + symbol depending on the charge of the residue. Point mutations are shown in red (negative) or blue (positive) solid circles emphasizing the charge change. Glycosylation acceptor sites are indicated with white (non-glycosylated) or black (glycosylated) dots.

B. To determine the topology *in vivo*, HEK-293T cells were transfected with Ct tagged (c-myc) E protein variants. The E protein virus and the proper mutations are indicated on top. Lanes with odd numbers are EndoH treated (+) and even numbers are mock treated (−). Bands of non-glycosylated, mono-glycosylated and double-glycosylated proteins are indicated by two white dots, one white and one black dots or two black dots, respectively.

References:

1. Bañó-Polo, M., Baldin, F., Tamborero, S., Marti-Renom, M. A. & Mingarro, I. N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* **20,** 179–186 (2011).
2. Tamborero, S., Vilar, M., Martínez-Gil, L., Johnson, A. E. & Mingarro, I. Membrane insertion and topology of the translocating chain-associating membrane protein (TRAM). *J Mol Biol* **406,** 571–582 (2011).
3. Surya, W., Li, Y. & Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1860,** 1309–1317 (2018).
4. Duart, G. *et al.* SARS-CoV-2 envelope protein topology in eukaryotic membranes. *Open Biol.* **10,** 200209–6 (2020).

4.3. Chapter 3: Interaction and function

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

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

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.

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

Programmed cell death is indispensable in multicellular organisms, contributing to the balance among cell death, proliferation, and differentiation that is crucial for tissue development and homeostasis[1]. Furthermore, protection from and defense against many disorders, including cancer, auto-immunity, and neurodegeneration, relies on apoptosis and autophagy[2,3]. Pathogen-related diseases are no exception, and efficient control and clearance of infections often require efficient programmed cell death[4,5]. Eliminating infected cells through a controlled death process blocks propagation of the infection and produces danger signals to stimulate an appropriate immune response[6].

Because of its relevance for cell fate, programmed cell death is tightly regulated. One of the primary modulators of apoptosis is the protein family known as B-cell lymphoma 2 (Bcl2)[7]. The Bcl2 family (consisting of ~20 proteins) incorporates pro-survival (e.g., Bcl2 and BclxL)[8], pro-apoptotic (e.g., Bax and Bak[9]), and BH3-only apoptosis activators (e.g., Bid and Bmf)[10]. Proteins are assigned to each subset based on their role in apoptosis (pro- vs anti-apoptotic), as well as on their sequence similarity to Bcl2. Most pro- and anti-apoptotic proteins share all four main Bcl2 sequence homology domains (BH1–4). In contrast, BH3-only members have only the BH3 domain, as their name implies. In addition, many Bcl2 family proteins, including some BH3-only members[11], have a transmembrane domain (TMD) in the carboxyl-terminal (Ct) end that effectively allows for insertion of the protein into the target lipid bilayer[12].

Cellular Bcl2 (cBcl2) proteins can physically interact with each other, forming homo- and hetero-oligomers[9,10,13–16]. These protein–protein interactions (PPIs) constitute an important regulatory mechanism of programmed cell death. In healthy cells, anti-apoptotic Bcl2s inhibit the activation of pro-apoptotic proteins, either through direct interaction or sequestering BH3-only activators[7]. Upon an apoptotic stimulus, BH3-only and pro-apoptotic proteins are discharged, promoting mitochondrial membrane permeabilization and the release of cytochrome c into the cytosol which, in turn, will activate the apoptosome. Interactions among Bcl2 family members have been thought to occur only through soluble domains, especially BH domains. However, recent findings suggested that Bax TMD interacts with pro-survival Bcl2 proteins[17], expanding the range of interactions involved in the control of programmed cell death.

Viruses have developed multiple strategies to modulate cell death, including masking of internal cellular sensors, caspase regulation, signaling cascade modulation, and mimicking of Bcl2 regulators[18,19]. Functional homologs of cBcl2, known as viral Bcl2s (vBcl2s; singular, vBcl2), are present in numerous viral families, including Herpesviridae, Poxviridae, Adenoviridae, and Birnaviridae[20]. The sequence homology between vBcl2 and cBcl2 proteins varies considerably. However, the crystal structures of some vBcl2s reveal a structural homology in key domains between cBcl2 and vBcl2 despite no sequence homology, suggesting functional commonalities[5,21].

Here, we show that the Ct hydrophobic regions of herpesviral and poxviral vBcl2s can operate as TMDs, effectively inserting the protein into its target membrane. The vBcl2 TMDs facilitate homo-oligomerization and hetero-oligomerization with pro- and anti-apoptotic cBcl2 TMDs inside the membrane and in the absence of apoptotic stimuli. These newly discovered interactions participate in the regulation of cBcl2, thus modulating programmed cell death upon apoptotic stimuli. Furthermore, our results suggest that inhibitors of these intramembrane PPIs could be used therapeutically against life-threatening viral infections characterized by short- and long-term deregulation of apoptosis.

Results

vBcl2 contain a functional Ct TMD. Like their cellular counterparts, many vBcl2s present a hydrophobic region in their Ct end. To identify whether these hydrophobic regions could act as true TMDs, we selected six vBcl2 proteins from two distinct viral families (3 herpesviruses and 3 poxviruses): BHRF1 (*Human gammaherpesvirus 4 – Epstein–Barr virus*, HHV4)[22], ORF16 (*Human gammaherpesvirus 8 – Kaposi's sarcoma–associated herpesvirus*, HHV8)[23], ORF16 (*Bovine gammaherpesvirus 4*, BoHV4)[24], F1L (*Vaccinia virus*, VacV)[25,26], M11L (*Myxoma virus*, MyxV)[26,27], and ORFV125 (*Orf virus*, OrfV)[28]. To avoid confusion, here we use the viral acronym to refer to the vBcl2 protein.

First, we analyzed in silico the presence of TMDs in the selected vBcl2. For this purpose, we used two TMD prediction software packages, TMHMM v2.0[29,30] and the ΔG prediction server[31,32]. Both algorithms identified a TMD in the Ct of HHV4, HHV8, MyxV, and OrfV. Although the ΔG prediction server identified a Ct TMD for BoHV4 and VacV, the TMHMM software did not (Supplementary Fig. 1).

We then aimed to explore the membrane insertion capacity of the predicted membrane-spanning segments using an in vitro assay based on the *E. coli* leader peptidase (Lep), an assay that allows for accurate and quantitative description of the membrane insertion capability of short sequences. The Lep sequence we employed contains an extended Nt luminal section, followed by two TMDs (H1 and H2) connected by a cytoplasmic loop, and a large luminal Ct domain. Two glycosylation acceptor sites were inserted into the Lep (G1 and G2), one in each luminal domain (Nt and Ct ends) (Fig. 1a). N-linked glycosylation occurs exclusively in the lumen of the endoplasmic reticulum (ER) because of the location of the oligosaccharyltransferase (OST) active site (a translocon-associated enzyme responsible for the oligosaccharide transfer[33]), thus serving as a topological marker. Glycosylation of an acceptor site increases the apparent molecular mass of the protein (~2.5 kDa), which facilitates its identification by gel electrophoresis. When H1 and H2 are present, both G1 and G2 acceptor sites locate in the ER lumen and are subsequently glycosylated. Likewise, if H2 is replaced by a sequence that the translocon efficiently recognizes as a TMD, a double (G1 and G2) glycosylation would be expected. In contrast, if H2 is deleted or substituted by a protein segment that the translocon does not recognized as a TMD, only G1 would remain in the lumen, and single glycosylation will be observed (Fig. 1a).

Using this system, we studied the insertion of all six vBcl2 hydrophobic Ct domains (Fig. 1b). Additionally, the single TMD of two mitochondrial membrane proteins, Tomm20 and Tomm22 (T20 and T22), together with a construct in which Lep was not altered were used as insertion controls. As a mono-glycosylated control, we used the Ct region of the BH3-only protein Noxa, a hydrophobic domain not recognized by the translocon as a TMD[11]. The sequences for all non-viral proteins referenced here can be found in Supplementary Fig. 2. Additionally, we used a Lep mutant from which H2 was deleted as a mono-glycosylated/non-insertion control (Supplementary Fig. 3). To ensure that the increase in molecular weight was due to glycosylation of the acceptor sites, OST samples were incubated in the presence (+) or absence (−) of Endoglycosydase H (EndoH), a glycan-removing enzyme. The results of the glycosylation assay indicated that all vBcl2 hydrophobic regions, including those derived from BoHV and VacV, were efficiently inserted into ER-derived microsomal membranes (Fig. 1c). Furthermore, the Lep assay revealed that all viral TMDs insert into ER-derived membranes more efficiently than predicted by the ΔG prediction server (Fig. 1b).
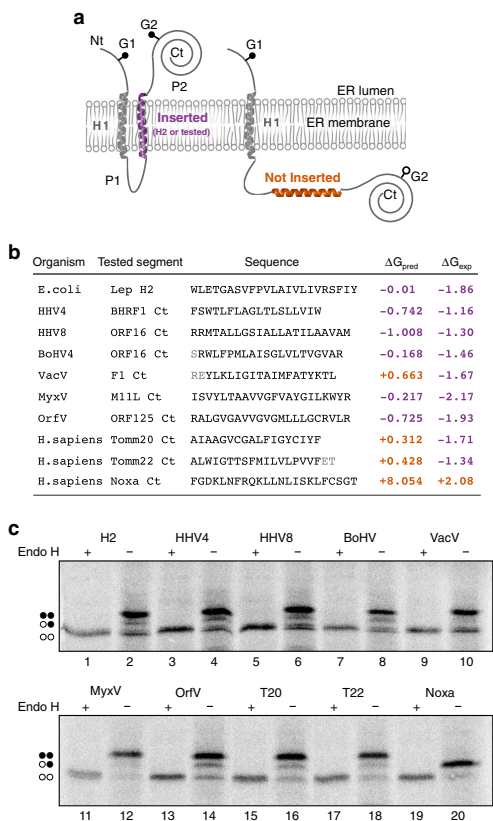
Fig. 1 Insertion of vBcl2 Ct hydrophobic sequences in eukaryotic membranes. a A schematic representation of the leader peptidase (Lep) model protein. G1 and G2 denote artificial glycosylation acceptor sites. The sequence under investigation was introduced replacing the H2 domain of Lep. Recognition of the tested sequence as a TMD by the translocon machinery (highlighted in purple) results in the modification of the G1 and G2 acceptor sites. The Lep chimera will be singly glycosylated if the sequence being tested is not recognized as a TMD and thus not inserted into ER-derived membranes (shown in orange). **b** The sequences (including name and organism of origin) used in the assay are shown alongside their predicted ΔG (ΔG_{pred}) in kcal/mol, calculated using the ΔG prediction server in its ΔG prediction mode. Residues shown in gray were not considered by the ΔG prediction server. Additionally, the experimental ΔGs in kcal/mol (ΔG_{exp}), calculated based on the ratios between mono and double glycosylated populations, are presented in the right column. Experimental results are the average of at least four independent experiments. Purple numbers indicate negative ΔGs (insertion of the tested sequence), while orange numbers denote ΔG values above 0. **c** A representative example ($n = 4$) of an in vitro protein translation in the presence of ER-derived microsomes and in the presence (+) or absence (−) of Endoglycosydase H (EndoH), a glycan-removing enzyme. The absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots.

Homo-oligomerization of the vBcl2 TMDs. cBcl2 TMD homo-oligomerization has been reported in biological membranes[17]. To assess whether vBcl2 TMDs exhibit this self-associating property, we employed bimolecular fluorescent complementation (BiFC) approach[34] adapted for the study of intramembrane interactions[17,35]. Briefly, TMDs were fused with two non-fluorescent fragments of the venus fluorescent protein (VFP), the VN (N terminus) and the VC (C terminus; Fig. 2a). Interaction of the TMDs facilitated the approximation of the VN and VC ends and the reconstitution of the VFP protein structure and, in consequence, the recovery of its fluorescent properties. As a positive control and normalization tool, we included the TMD of glycophorin A (GpA), a sialoglycoprotein found in human erythrocyte membranes that can form non-covalent dimers solely through the association of its single hydrophobic TMD[36–38]. The non-oligomerizing TMDs of the mitochondrial membrane proteins T20 and T22 and H2 from Lep were used as negative controls for membrane overcrowding (Fig. 2b).

The TMDs of HHV4, HHV8, VacV, MyxV, and OrfV showed an interaction capability similar to the controls, including T20 (used for the statistical analysis) and similar to that observed with Bcl2 TMD (Fig. 2b). However, BoHV did not show VFP fluorescence values significantly higher than the T20 control (Fig. 2b). Western blot analysis showed comparable expression levels of all chimeras (Supplementary Fig. 4). Because of the large variability, we decided to revise the results obtained with BoHV TMDs. For this purpose, we used BlaTM, a genetic tool designed to study TMD–TMD interactions in bacterial membrane[39].

Briefly, to measure interactions between TMDs, the designed sequences were fused to either the Nt or the Ct end of a split β-lactamase (βN and βC, respectively) and to the green fluorescent protein (GFP) (Supplementary Fig. 5a). These chimeras also included the pelB cleavable signal peptide, which directs the protein to the inner bacterial membrane and determines its topology, ensuring a periplasmic localization of the β-lactamase. Once located in the inner membrane, an efficient TMD–TMD interaction facilitates the reconstitution of the β-lactamase and thus the growth of the bacteria in selective media (ampicillin). In this assay, the LD_{50} of the antibiotic served as an indicator of the interaction strength (Fig. 2c, d), and GFP fluorescence allowed rapid quantification of protein levels (Supplementary Fig. 5b). As in the previous experiments, the TMD of GpA was used as a positive control and normalization value across experimental replicas, and the TMD of T20 was used as a negative control. Additionally, the HHV8 TMD was included in the assay as an extra control. In the BlaTM assay, both BoHV TMD and HHV8 TMD showed an interaction (ampicillin LD_{50}) above the negative control (Fig. 2c, d). Expression levels (measured using the GFP fluorescence) were comparable for GpA, T20, HHV8, and BoHV (Supplementary Fig. 5b).

By directly localizing to the mitochondria, Bcl2 controls mitochondrial outer membrane (MOM) permeabilization and, subsequently, apoptosis. In the mitochondria, Bcl2 can interact through its TMD with other members of the Bcl2 family[17]. Nonetheless, Bcl2 also localizes to other intracellular compartments, including the ER and the Golgi apparatus[40]. To analyze the location of the observed vBcl2 TMD homo-oligomers, we expressed the appropriated BiFC partner combinations together with an ER, mitochondrial, or plasma membrane marker (Supplementary Fig. 6). Our results showed that all the assayed TMD homo-oligomers occurred at the ER and at the mitochondrial membranes. Of interest, the TMD oligomers of VacV and MyxV could also be observed in the plasma membrane. These results suggest a similar distribution for vBcl2 and cBcl2.
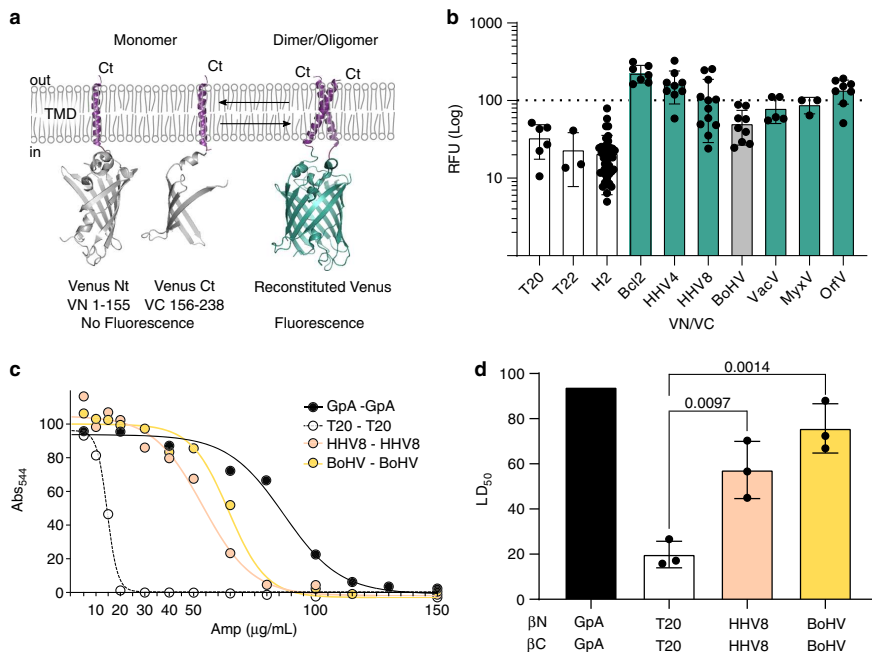
Fig. 2 Homo-oligomerization in biological membranes. a A schematic representation of the Bimolecular Fluorescent Complementation (BiFC) assay. The position of the Nt and Ct ends and the TMDs included in the constructs are indicated. The residues included in each VFP fragment are indicated below the protein representation. **b** Relative fluorescence units (RFU) of each tested homo-oligomer in the BiFC assay. The mean and standard deviation of at least three independent experiments are represented (n values 6, 3, 37, 9, 9, 12, 9, 5, 3, 9, respectively). The individual value of each experiment is represented by a solid dot. The VN-GpA/VC-GpA homodimer, used as a positive control and to normalize values across experiments, is represented by a dotted line. The T20, T22, and Lep H2 TMDs were included as negative controls (white bars). Those homo-oligomer that produced fluorescence levels significantly higher than the T20 TMD homo-oligomer (two-tailed homoscedastic t-test) are highlighted in green, and those that did not are shown in light gray. **c** Representative examples of dose–response curves used to calculate LD_{50} values in the BlaTM assay. The TMD homodimer of GpA was used as a positive control (black), while the T20 TMD was used as a negative control (white). **d** BlaTM calculated mean and standard deviation LD_{50} of at least three independent experiments (n = 3). The individual value of each experiment is represented by a solid dot. The βN-GpA/βC-GpA homodimer was used as a positive control (black bar). The level of significance (p-value, two-tailed homoscedastic t-test) when comparing T20 (white bar) vs HHV8 (pink bar) or BoHV (yellow bar) homo-oligomers is shown above the bars.

**cBcl2 and vBcl2 TMDs hetero-oligomerize in eukaryotic membranes**. Next, we decided to investigate the potential TMD–TMD interactions between vBcl2s and cBcl2s. For this purpose, we used the previously described BiFC approach. For any given interaction (i.e., X/Y), two possible combinations with the VFP fragments exist (VN-X/VC-Y and VN-Y/VC-X), so we investigated both. Once again, the GpA TMD homodimer was used as a positive control and a normalization value. As negative controls, we used the interaction of each partner in the hetero-oligomers with T20 TMD (i.e., for the X/Y interaction, X/T20 and T20/Y). An interaction was considered only if the fluorescence values were significantly higher than those of the two particular negative controls. To reinforce our data, we included in the study the interactions between the vBcl2 TMDs and T22 or H2 TMDs. The mean relative fluorescence and standard deviation of all the controls and the interactions between vBcl2 TMD and T22 TMD and Lep H2 can be found in Supplementary Fig. 7. The assay was carried out in human-derived cells despite the natural host of the virus; however, the amino acid composition of the cBcl2s TMDs are well conserved across the corresponding species (Supplementary Fig. 8). First, we investigated the potential TMD–TMD interactions between vBcl2s and the anti-apoptotic cBcl2s Bcl2 and BclXL (Fig. 3a, b). All of the viral TMDs could interact with the TMD of Bcl2 in at least one of the assayed combinations. On the other hand, although the majority of vBcl2 TMDs also interacted with BclXL TMDs, BoHV and MyxV TMDs did not (regardless of the partner combination used in the screening).

Next, because of the strength and consistency of the observed signal, we decided to further characterize the Bcl2-HHV8 TMD–TMD interaction. First, we corroborated this intramembrane interaction by performing a competition assay. The Bcl2 and HHV8 TMD-driven homo-oligomers were challenged either with the Bcl2 full-length protein (Bcl2-FL) or with a Bcl2 mutant lacking the TMD (Bcl2 ΔTMD). The Bcl2 TMD homo-oligomer was hindered when Bcl2-FL was co-expressed but not when the Bcl2 ΔTMD construct was included in the assay (Fig. 4a). Similarly, the addition of Bcl2-FL to the HHV8 TMD homo-oligomer decreased the observed fluorescence compared with the effect of the presence of Bcl2 ΔTMD. This result confirmed that HHV8 and Bcl2 TMDs can interact and suggested that the soluble domain of Bcl2 does not preclude an interaction between TMDs.
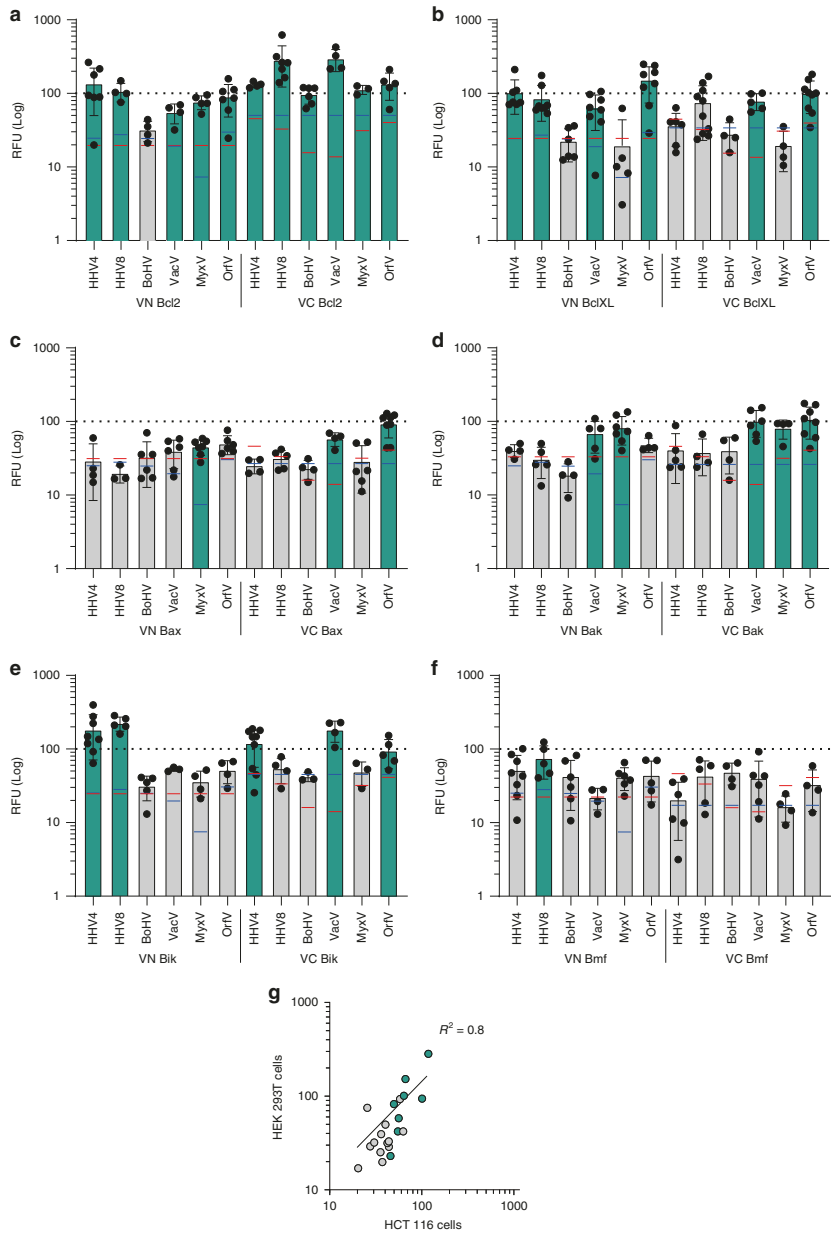
Fig. 3 Hetero-oligomerization in Eukaryotic membranes. a–f Relative fluorescence (RFU) for the hetero-oligomerization of vBcl2 and cBcl2 TMDs. The mean and standard deviation of at least three independent experiments are shown. Solid dots represent the results of individual experiments (n ranges between 3 and 10). The TMD included in the VFP chimeras (VN or VC) is indicated below each bar. The GpA TMD homodimer was used as a positive control and as the normalization value (dotted line). The interactions of each partner in the hetero-oligomers with T20 TMD were used as negative controls (i.e., X/T20 and T20/Y for the X/Y interaction). The blue (VN/T20 TMD) and red (VC/T20 TMD) lines within each bar indicate the fluorescence of the controls. An interaction (highlighted in green) was considered only if the RFU obtained was significantly higher (two-tailed homoscedastic t-test, p-value < 0.05) than those of both negative controls. **g** Correlation of the BiFC assay results in HCT 116 and HEK 293T cells. Green dots indicate interactions in HEK 293T cells. A linear trend line and the corresponding coefficient of correlation are shown.
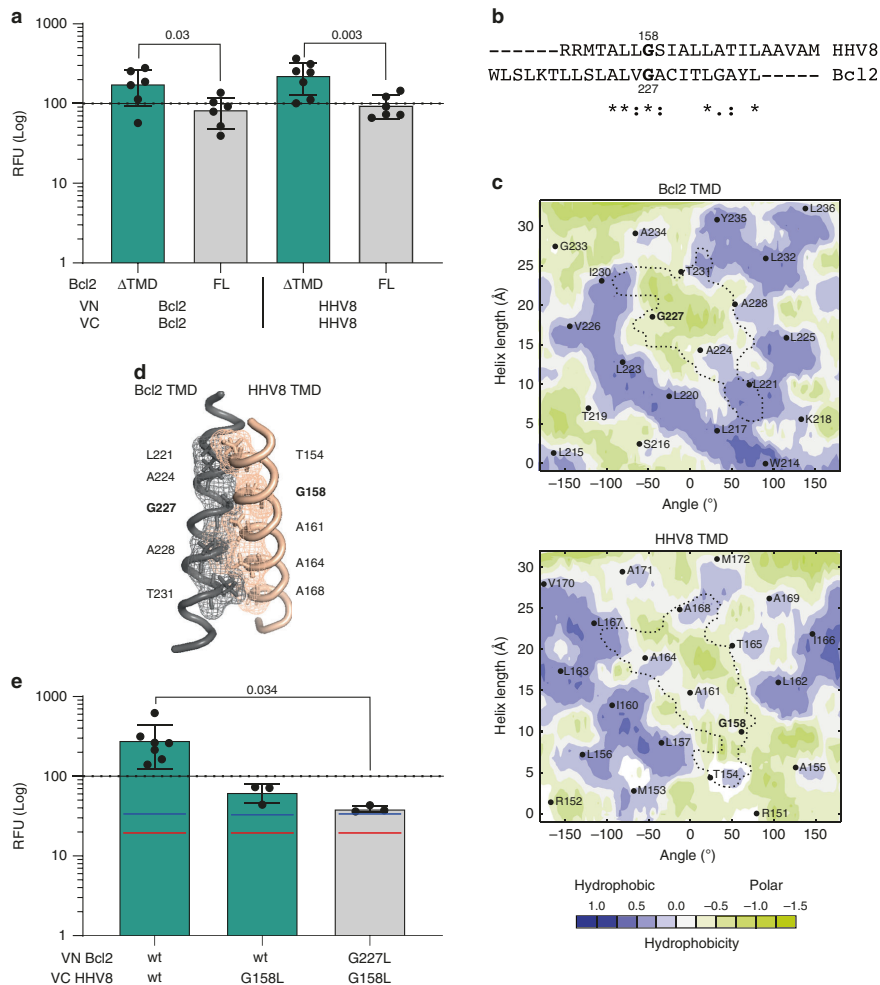
Fig. 4 Interaction between HHV8 and Bcl2 TMDs. **a** The BiFC-derived fluorescence (RFU) for the homo-oligomerization of Bcl2 (left) or HHV8 (right) TMDs was measured in the presence of Bcl2 (FL) or Bcl2 ΔTMD (ΔTMD) in HEK 293T cells. The mean and standard deviation of at least six independent experiments are shown ($n \geq 6$). Solid dots represent the results of individual experiments. The GpA TMD homodimer was used as a positive control and as the normalization value (dotted line). A significant (two-tailed homoscedastic t-test) decrease in the fluorescence is highlighted in gray, p-values are indicated. **b** Sequence alignment of Bcl2 and HHV8 TMDs. The alignment was done with Clustal Omega (EMBL-EBI). **c** Predicted (PredDIMER) contact area between Bcl2 and HHV8 TMDs. The position (length and angle) of each residue in a putative α-helix are indicated. A dotted line encircles the contact area in each TMD. The hydrophobicity of the residues is shown with a color scale. As expected, there are many hydrophobic residues in both TMDs. However, the contact area between both α-helices is mostly hydrophilic. **d** Model of a putative dimer between Bcl2 TMD and HHV8 TMD, obtained with PredDIMER. The residues involved in the interaction are indicated, conserved glycine residues in Bcl2 and HHV8 are highlighted in bold. **e** Bcl2 and HHV8 TMD hetero-oligomerization measured by BiFC in HEK 293T cells. Fluorescence from the VN GpA/VC-GpA homodimer was used to normalize the signal (dotted line). The mean and standard deviation of at least three independent experiments are shown ($n \geq 3$). Solid dots represent the results of individual experiments. The blue (VN T20 TMD) and red (VC T20 TMD) lines indicate the fluorescence of the controls used for the statistical analysis (two-tailed homoscedastic t-test), green bars denote RFU values above the controls, p-values are indicated.

An in silico analysis of the potential dimer interface (done with Preddimer, a TM segment dimer prediction software program[41]) suggested that conserved glycine residues between Bcl2 and HHV8 TMDs (at positions 158 and 227, respectively) could be located at the contact surface of a putative HHV8-Bcl2 TMD–TMD heterodimer (Fig. 4b–d, and Supplementary Fig. 9a). This positioning would create the appropriate environment for ridge–groove arrangements, as observed for other interacting TMDs[36,37,42]. Single amino acid substitutions were designed to analyze the role of these glycine residues. Precisely, a mutation of

glycine 158 to leucine in HHV8 TMD (G158L) reduced the observed interaction between Bcl2 and HHV8 TMDs in the BiFC assay (Fig. 4e). Moreover, mutation of glycine residues in both helices (G158L in HHV8 and G227L in Bcl2) further decreased the fluorescence signal to levels significantly below those obtained with the wild-type forms of the TMDs. Of note, the introduction of G158L in HHV8 or G227L in Bcl2 did not alter insertion into biological membranes (Supplementary Fig. 3). These results indicate that the TMDs of Bcl2 and HHV8 can interact efficiently and specifically at the mitochondrial and ER membranes through arrangements in which the glycine residues have a crucial role.

We next analyzed the interactions between vBcl2 TMDs and the TMDs of the cellular pro-apoptotic Bax and Bak proteins (Fig. 3c, d). All three poxviral TMDs (VacV, MyxV, and OrfV) could interact with Bax and Bak TMDs in the absence of apoptotic stimuli and independent of any cytosolic (soluble) domain interactions. On the other hand, herpesviral HHV4, HHV8, and BoHV TMDs showed no interaction with either Bax or Bak TMDs. MyxV protein interacts with Bak and Bax[43,44] inhibiting the conformational activation of Bax[43]. Based on this information, we decided to investigate the role of MyxV TMD and its interaction with Bax TMD in the viral protein function. For this purpose, we first decided to confirm the host–pathogen intramembrane interaction using the previously described BlaTM assay. The MyxV and Bax TMDs were inserted into the βN and βC chimeras, respectively, and assayed together or in combination with the complementary chimeras carrying the TMD of T20 (Fig. 5a, b). The results of the BlaTM assay confirmed the interaction between MyxV and Bax TMDs, which showed LD_{50} values that were significantly higher than the βN-T20/βC-T20, βN-MyxV/βC-T20, and βN-T20/βC-Bax combinations.

Using the Preddimer algorithm, we analyzed the Bax TMD–MyxV TMD interaction (Fig. 5c, d and Supplementary Fig. 9b). Once again, the model showed an interaction in which a ridge–groove arrangement was created by an adequate disposition of large and small residues. Of interest, a large proportion of aromatic residues was found at the interface of these two TMDs. Amino acid substitutions were again designed to analyze the role of the small residues. Mutation of glycine 158 to isoleucine in MyxV TMD (G158I) or double mutation of glycine 179 and alanine 183 to isoleucine in Bax TMD (G179I A183I) reduced the interaction of Bax and MyxV TMDs in the BlaTM assay (Fig. 5e). Moreover, the introduction of mutations in both helices (G158I in MyxV and G179I A183I in Bax) further decreased antibiotic resistance. Of note, mutations G158I in MyxV and G179I A183I in Bax TMD did not alter membrane insertion (Supplementary Fig. 3).

In eukaryotic cells, ectopic expression of Bax TMD induces some caspase 3/7 activation and subsequently cell death, which expression of the anti-apoptotic Bcl2 TMD can counteract, thanks to the interaction among Bcl2 and Bax TMDs[17]. We wondered whether the interaction between MyxV and Bax TMDs also could preclude Bax TMD caspase activation. To test this hypothesis, we co-expressed H2, Bcl2, or MyxV TMD with Bax TMD and measured the resulting cell viability (Fig. 5f). In agreement with the hetero-oligomerization results, Bcl2 and MyxV TMDs, but not control H2 TMD, interfered with the Bax TMD-induced apoptosis.

Next, to extend the BiFC-based screening of the intramembrane host–pathogen interactions between vBcl2s and cBcl2s, we studied the interactions with the BH3-only apoptotic modulators Bik and Bmf (Fig. 3e, f). The TMDs of HHV4, HHV8, VacV, and OrfV could interact with the Bik TMD. However, according to our results, viral interactions with Bmf were much sparser, and only the HHV8 TMD could interact with Bmf TMD. All intramembrane PPIs found between cellular and viral proteins are

summarized in Supplementary Fig. 10. For a more comprehensive visualization, we also have included a network representation (Supplementary Fig. 11). Additionally, the result of the BiFC-based HEK 293T screening was confirmed in HCT 116 cells. Despite some differences, a strong correlation between the BiFC assay in both cell types can be observed ($R^2 = 0.8$) (Fig. 3g and Supplementary Fig. 12).

While studying interaction with the BiFC assay (i.e., X/Y), we observed some differences in the results obtained with each of the two possible combinations (VN-X/VC-Y and VN-Y/VC-X).

Due to the nature of the BiFC assay, it is possible, as has occurred in the case of the BlaTM assay, that the reporter signal resulting from a TMD–TMD interaction depends not only on the sequence of the interacting TMDs, and thus its inherent affinity, but also on the orientation of the interacting surfaces of the TMDs in relation to the accompanying signaling domains[39]. To test our hypothesis we chose the first of the interactions where significant differences between the two possible BiFC combinations were found, i.e., the interaction between Bcl2 and BoHV TMDs (Fig. 3a). Our screening revealed VN-BoHV/VC-Bcl2 as a TMD–TMD interaction but not VN-Bcl2/VC-BoHV. To change the orientation of the TMD with respect to the reporter domain we inserted one (+1), two (+2), or three (+3) native residues at the Nt end of BoHV TMD[39]. The observed fluorescence for the interaction of the new BoHV TMD variants in combination with the Bcl2 TMD were higher (and significantly above their controls) than the interaction with the original design of the VN-BoHV chimera (Supplementary Fig. 13). Although this result confirms our hypothesis, due to a large number of interactions studied and despite a large number of replicas, we cannot rule out the presence of some false positive or negative results.

Collectively, these results point to an intricate network of interactions between the TMDs of viral and cellular origin. A comprehensive analysis of the data using a principal component analysis (PCA) revealed that HHV4 and HHV8 TMDs from herpesviral proteins behave similarly, as would have been expected given their taxonomic proximity (Fig. 6a, b). Likewise, the TMDs of a poxviral origin, particularly VacV and OrfV TMDs, were tightly grouped in the PCA. However, the observed similarities could not have been inferred by the analysis of the TMDs sequences (Fig. 6c), which suggests a structural pattern underlying the sequence that governs the TMD interactions.

Viral–host TMD–TMD interactions are necessary to modulate apoptosis. To identify whether these newly found viral–host TMD–TMD interactions are necessary to control cellular apoptosis, we transfected HeLa cells with Bcl2, HHV8, or MyxV either with or without the TMD (FL and ΔTMD, respectively). Additionally, we included chimeras in which the TMD of each of the previously mentioned proteins was replaced by the TMD of T20 (Bcl2-T20, HHV8-T20, and MyxV-T20, respectively), which our previous results indicated cannot interact with any cBcl2 TMD. Once transfected, cells were treated with doxorubicin to induce apoptosis[45].

As expected, the FL proteins could prevent apoptosis (Fig. 7a, and Supplementary Fig. 14). However, once the TMD was removed, none of the proteins could promote survival or stop apoptosis. Similarly, the chimeras carrying the TMD of T20 could not control the doxorubicin-induced apoptosis, suggesting that the function of the viral TMDs extends beyond anchoring the protein to the membrane and that TMD–TMD hetero-oligomerizations are crucial for modulating viral-induced cell death (Fig. 7a, b). Western blot analysis confirmed that differences in expression levels were not the source of the observed results (Fig. 7c).
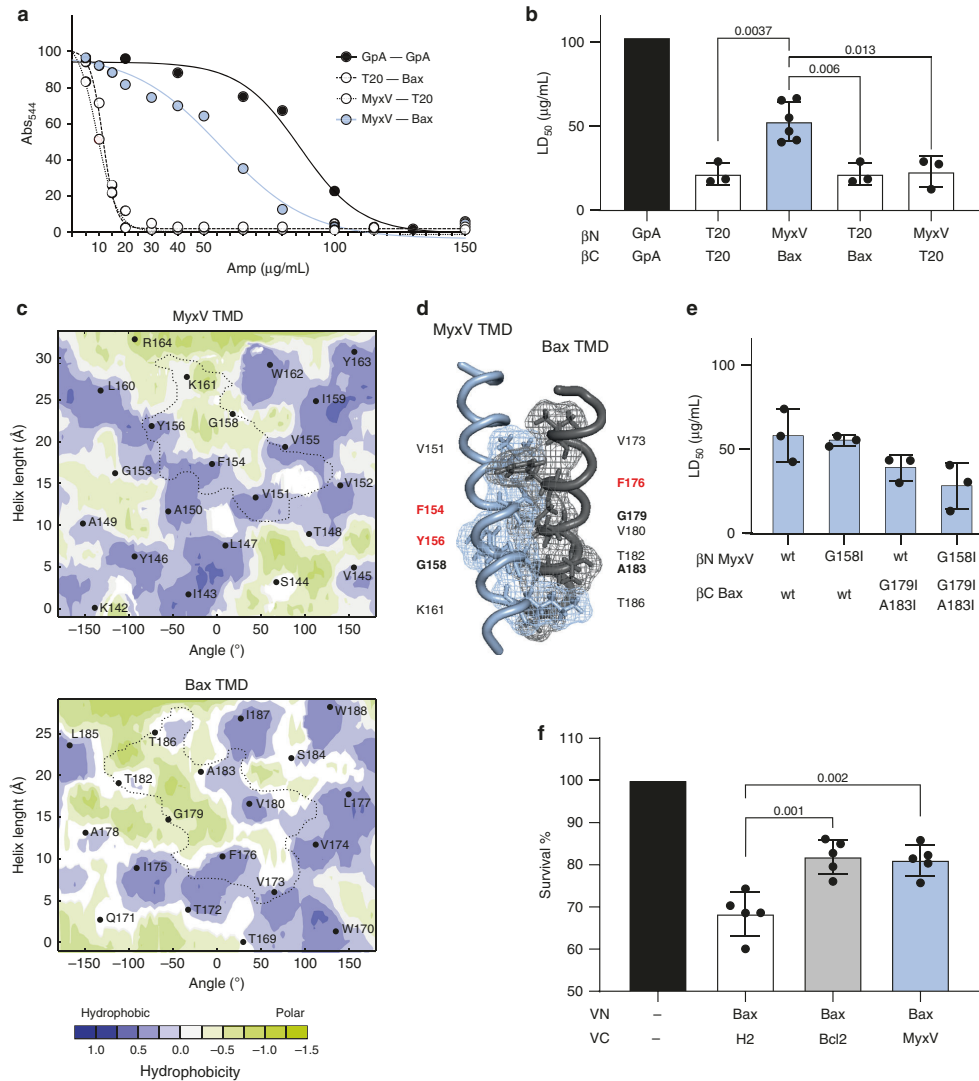
Fig. 5 Interaction between MyxV and Bax TMDs. a Representative example of dose–response curves used to calculate LD_{50} values in the BlaTM assay. The TMD homodimer of GpA was used as a positive control (black), while the T20 TMD was used as a negative control (white). **b** Summary of the results for the BlaTM assay. The indicated βN and βC chimeras were expressed in *E. coli* and the resulting ampicillin LD_{50} measured. LD_{50} mean and standard deviation of at least three independent experiments ($n \geq 3$) are shown. Two-tailed homoscedastic *t*-test *p*-values are indicated. Solid dots represent the results of individual experiments. **c** Predicted (PredDIMER) contact area between MyxV and Bax TMDs. The position (length and angle) of each residue in a putative α-helix is indicated. A dotted line encircles the contact area in each TMD. The hydrophobicity of the residues is shown with a color scale. **d** Model of a putative dimer between MyxV TMD and Bax TMD, obtained with PredDIMER. The residues involved in the interaction are indicated. Glycine and alanine residues are highlighted in bold while aromatic residues are shown in red. **e** BlaTM analysis for the interaction between MyxV and Bax TMDs carrying substitutions of Gly 158 (MyxV) and/or Gly 179 and Ala 183 (Bax) by Ile. LD_{50} mean and standard deviation of at least three independent experiments are shown ($n \geq 3$), solid dots represent the results of individual experiments. **f** Bax TMD-induced cell death. Cells were transfected with Bax TMD and H2, Bcl2 or MyxV TMD or mock-transfected; 24 hpt cell viability was measured by Trypan blue staining. Survival percentage mean and standard deviation of five independent experiments ($n = 5$) are shown. Two-tailed homoscedastic *t*-test *p*-values are indicated. Solid dots represent the results of individual experiments.
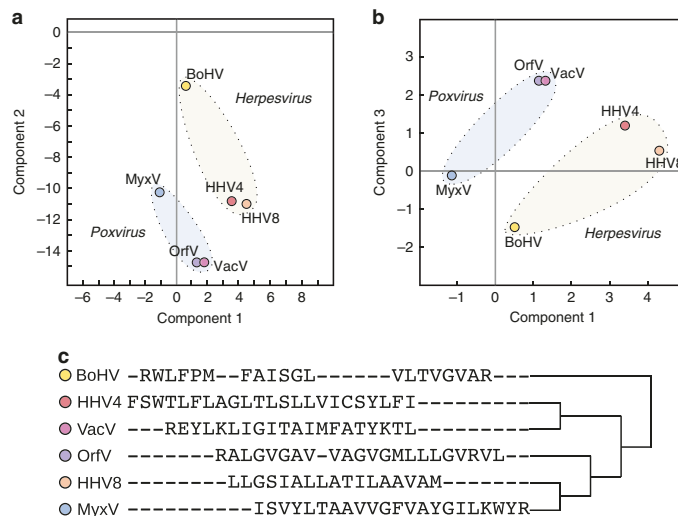
Fig. 6 TMD interaction and sequence profile. a, b Principal component analysis (PCA) of vBcl2 TMDs interaction profile with host TMDs. The colored areas highlight the viral families, with poxviruses in light blue and herpesviruses in light yellow. The analysis was performed using only the BiFC assay results. **c** vBcl2 TMDs Multiple Sequence Alignment. Results obtained with Clustal Omega (EMBL-EBI) using the default parameters.

Despite T20 being a mitochondrial-localized protein, the substitution of the natural TMD on Bcl2, HHV8, or MyxV by the T20 TMD could affect the localization of these proteins and impede their anti-apoptotic role. To investigate this possibility, we compared the localization of all three protein variants (FL, ΔTMD, and T20) by fluorescent confocal microscopy. The confocal analysis revealed that substitution of the natural TMD with the T20 TMD did not alter the localization of the protein (Supplementary Fig. 15). Furthermore, colocalization with a mitochondrial marker was observed with the Bcl2, HHV8, and MyxV-FL and all T20 variants. In contrast, the elimination of the Ct TMD generated soluble proteins in all three cases. These results suggest that the TMD, present in HHV8 and MyxV, not only facilitated proper localization of the proteins but is also necessary to establish interactions with host TMDs, which are essential for the viral control of apoptosis.

Next, we transfected HeLa cells with Bcl2 and MyxV (FL and T20 variants for both proteins) and induced apoptosis with Bax-FL (Fig. 7d). Both Bcl2 and MyxV-FL proteins could rescue Bax-induced apoptosis. However, the substitution of the TMD in either Bcl2 or MyxV by the Ct membrane anchoring segment of T20 eliminated their anti-apoptotic properties. Collectively, these results not only suggest a potential role for the Bax–MyxV TMD–TMD interaction during the viral infection but also reinforce our screening results.

It has been proposed that MyxV helices 2, 3, 4, and 5 form a binding groove for Bak and Bax[46]. Mutations in key positions of this groove alter MyxV binding to the aforementioned pro-apoptotic cellular proteins and subsequently disrupt its anti-apoptotic action. To compare the influence of in-groove vs TMD interactions, HeLa cells were transfected with MyxV-FL, T20, or MyxV bearing substitutions in alanines 71 or 82 to phenylalanines (A71F and A82F respectively)[46]. To induce apoptosis, cells were co-transfected with Bax-FL. Our results confirmed that mutations in the binding groove of MyxV, primarily A82F, alter MyxV pro-survival function (Supplementary Fig. 16a). However, the substitution of MyxV TMD by T20 TMD produced a stronger

effect on the ability of the protein to rescue Bax-induced apoptosis. Western blot analysis confirmed comparable expression levels for all of the FL, T20 A71F, and A82F variants (Supplementary Fig. 16b).

vBcl2 and cBcl2 TMDs interactions control viral-induced apoptosis. Finally, we aimed to explore whether vBcl2 TMD interactions are necessary to control the apoptosis induced by a viral infection. For this purpose, we infected HeLa cells with VacV-PKR, a VacV recombinant strain expressing the double-stranded RNA-dependent protein kinase (PKR) and capable of inducing apoptosis[47] and transfected with plasmids carrying Bcl2, Bcl2-ΔTMD, Bcl2-T20, HHV8, HHV8-T20, MyxV, or MyxV-T20 (under the control of an IPTG inducible promoter) or mock-transfected with an empty plasmid. At 24 h.p.i., levels of caspases 3 and 7 activity and cell death were measured as a readout for apoptosis (Fig. 8a, b). Bcl2-FL protein could partially block the PKR-induced apoptosis[48]. However, Bcl2-ΔTMD or Bcl2-T20 did not rescue cell death. Similarly, HHV8-FL and MyxV-FL significantly reduced caspase 3/7 levels and decreased cell death. On the other hand, HHV8-T20 or MyxV-T20, which includes a TMD incapable of establishing intramembrane interactions, did not reduce caspase 3/7 levels and were unable to control cell death. To confirm these results, we infected HeLa cells with a *Modified Vaccinia Ankara virus* strain lacking F1L protein (MVA-ΔF1L) and capable of inducing apoptosis[49] and transfected with Bcl2-FL, Bcl2-ΔTMD, Bcl2-T20, HHV8-FL, HHV8-T20, MyxV-FL, MyxV-T20, VacV-FL, pr VacV-T20 (under the control of IPTG) or mock-transfected with an empty plasmid (Fig. 8c, d). Additionally, the use of MVA-ΔF1L allowed us to explore the role of TMD interactions in VacV vBcl2 (F1L). As in the previous experiment (VacV-PKR, panels a and b), only proteins bearing a TMD capable of establishing interactions (FL variants) were able to control cellular apoptosis. In both experiments, the western blot analysis confirmed comparable expression levels for all the FL, ΔTM, and T20 constructs (Fig. 8e). In summary, our results demonstrate that interactions between viral
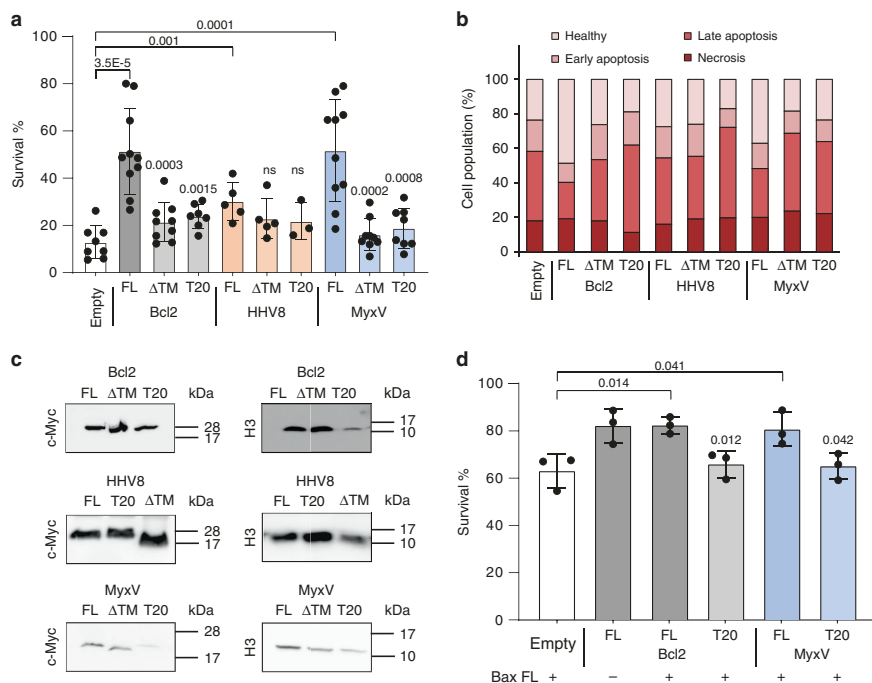
Fig. 7 The role of vBcl2 TMD in apoptotic control. a Role of vBcl2 TMD in control of doxorubicin-induced apoptosis. Cells were transfected with Bcl2, HHV8, and MyxV with (FL) or without the TMD (ΔTMD), or with the TMD substituted by the TMD of T20 (T20). Cells were then treated with doxorubicin and the percentage of surviving cells was calculated by Trypan blue staining. The survival percentage mean and standard deviation of at least three independent experiments are shown ($n \geq 3$). Solid dots represent the results of individual experiments. Transfection with an empty plasmid (Empty) was used as a negative control. Statistical differences are based on a two-tailed homoscedastic t-test (p-values are indicated above the corresponding bar, ns non-significant). **b** Alternatively, the percentage of cells in healthy, early apoptotic, late apoptotic, or necrotic states was measured by flow cytometry using propidium iodide staining and phosphatidyl serine labeling (FITC-Annexin V). The average percentage of cells in each of the aforementioned populations is shown. **c** Western blot analysis of protein levels. Histone 3 (H3) was used as a loading control ($n = 3$). **d** Role of MyxV TMD in Bax-induced apoptosis. Cells were transfected with Bcl2 and MyxV (FL) or with the TMD substituted by the TMD of T20 (T20) and co-transfected with (+) or without (−) Bax-FL as an apoptosis stimulus. Survival percentage mean and standard deviation of three independent experiments are shown ($n = 3$). Solid dots represent the results of individual experiments. Statistical differences are based on a two-tailed homoscedastic t-test (p-values are shown above the corresponding bars).

and host TMDs in the outer mitochondrial membrane are required for the control of apoptosis in a viral infection scenario.

Discussion

To prevent the premature death of host cells, many viruses have acquired the ability to modulate apoptosis. As a result, the virus can replicate longer and circumvent activation of the immune system[18,50]. This infection-induced effect can be permanent, provoking the cell to persistently escape programmed cell death despite paracrine or autocrine stimulus[51]. Indeed, 8–16% of new cancer cases are attributable to carcinogenic infections[52,53]. Preventing acute infection episodes as well as the long-lasting effects associated with apoptosis evasion (including cancer) requires a detailed understanding of how viruses control programmed cell death. Viruses employ multiple strategies to block apoptosis. One of the most notable, at least in DNA viruses, is the expression of cellular pro-survival Bcl2 analogs[20]. Remarkably, many structural elements of cBcl2, including a Ct hydrophobic region, are preserved in their viral replicas. Here, we explored the role of the vBcl2 Ct domain in protein function. More specifically, we

focused on its function as a membrane anchor, its contribution to host–host and viral–host PPIs, and its importance in the control of cellular apoptosis.

In this report, we demonstrated that the cellular machinery recognizes Ct regions of the tested vBcl2s as truly being TMDs and therefore inserts them into the membrane despite a low hydrophobicity score for some of these regions (BoHV, MyxV). Furthermore, a positive (suggesting no insertion) ΔG_{pred} value for the insertion into lipid bilayers was predicted for VacV Ct region. Nonetheless, this predicted penalty did not preclude VacV Ct insertion. In fact, all the ΔG_{exp} values obtained in vitro were below −1 kcal/mol, suggesting a strong membrane insertion capacity.

Permeabilization of the MOM, considered in most cases the point of irreversible commitment to apoptosis, is controlled by interactions among pro- and anti-apoptotic Bcl2 members[54]. Most members of the Bcl2 family, including BH3-only proteins, bear a Ct TMD necessary for both their subcellular localization and apoptotic activity. Additionally, Bax and Bcl2 TMDs have previously been found to interact in biological membranes[17]. This
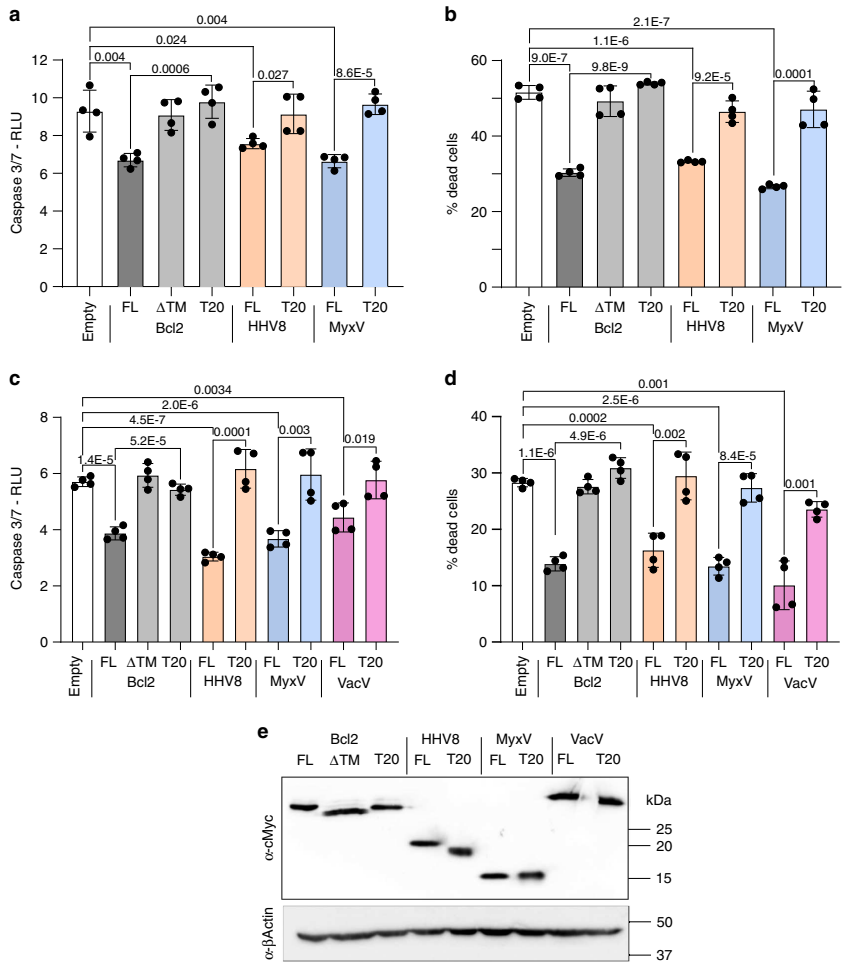
Fig. 8 The role of viral TMDs on viral-induced apoptosis control. a, b HeLa cells were co-infected with VacV-PKR and VT7LacOI (MOI 2.5 for each virus) and transfected with 50 ng of the plasmids Bcl2-FL, Bcl2-ΔTMD, Bcl2-T20, HHV8, HHV8-T20, MyxV, and MyxV-T20 or with an empty plasmid. IPTG was then added to induce protein expression. At 24 h.p.i, apoptosis was measured from duplicate samples using the Caspase-Glo 3/7 assay kit (**a**) or by measuring cell viability with Trypan blue staining (**b**). The mean and standard deviation of four technical replicates from two independent experiments are shown. Solid dots represent the results of individual replicates. Cells transfected with an empty plasmid were used as a control. Statistical differences are based on a two-tailed homoscedastic t-test, p-values are indicated above the corresponding bar. **c, d** HeLa cells were co-infected with VT7LacOI and MVA-C-ΔF1L deletion mutant (MOI 2.5 of each virus) and transfected with 50 ng of the plasmids Bcl2-FL, Bcl2-ΔTMD, Bcl2-T20, HHV8, HHV8-T20, MyxV, MyxV-T20, VacV, and VacV-T20 or with an empty plasmid. IPTG was then added to induce protein expression. At 24 h.p.i, apoptosis was measured from samples using the Caspase-Glo 3/7 assay kit (**c**) or by measuring cell viability by Trypan blue staining (**d**). The mean and standard deviation of four technical replicates from two independent experiments are shown. Cells transfected with an empty plasmid were used as control. Statistical differences are based on a two-tailed homoscedastic t-test, p-values are indicated above the corresponding bar. **e** Western blot analysis of Bcl2, HHV8, MyxV, and VacV with (FL) or without the TMD (ΔTMD) or with the TMD of T20. All viral proteins included a c-Myc tag. Beta-actin was used as a loading control ($n = 3$).

information, coupled with the newly discovered membrane-spanning capacity of the vBcl2 Ct hydrophobic region, led us to explore the homo-oligomerization potential of the newly identified vBcl2 TMDs. Our findings using two different complementation assays, i.e., BiFC and BlaTM, suggest that all vBcl2 TMDs can form intramembrane homo-oligomers. Sequence-specific TMD oligomerization is influenced by lipid bilayer properties and thus its composition[55–57]. Our results indicate that the BiFC-observed homo-oligomers can form in vivo at the ER and mitochondrial membranes. It is particularly interesting to find these anti-apoptotic homo-oligomers at the mitochondrial membrane, the site of the pro-apoptotic action of Bax, Bak, and

other pro-apoptotic Bcl2 proteins[58]. These results not only indicate a similar distribution for vBcl2 and cBcl2 but also raise the possibility of a potential dynamic intracellular localization for vBcl2.

Despite some drawbacks, such as an elevated number of replicates needed or difficulty automating the process, we believe that the BiFC-based approach is particularly well suited for our study. The BiFC assay not only allowed us to conduct our experiments in eukaryotic cells but also replicated the native membrane topology of the vBcl2 proteins. Proteins with no signal peptide and a single TMD in their Ct end (known as tail-anchored proteins), such as the Bcl2 family, are a minority within the cellular membrane proteome (3–5%)[59]. By locating the TMD at the Ct end of our chimeras, we replicated its natural location/orientation and any potential for TMD–TMD interaction. Furthermore, the BiFC assay facilitated the study of hetero-oligomerizations and the identification of the subcellular compartment where the interactions occur.

Using BiFC, we also analyzed the potential TMD–TMD interactions between vBcl2 and cBcl2. Our results revealed a previously overlooked host–pathogen intramembrane interaction network. It should be noted that intramembrane oligomers can be fragile because of the small interaction surface between monomers[44,60–62]. This history indicates that at least in some cases, the contact between the cellular and viral proteins occurs at the same time through soluble and membrane-spanning regions.

Of interest, all the viral TMDs we tested here, except that of BoHV, could interact with multiple cellular TMDs. The particularities of these connection circuits varied from virus to virus, potentially reflecting distinctive mechanisms of action. Indeed, a comprehensive PCA of the TMD–TMD interaction data revealed a similar interaction pattern for herpesviral HHV4 and HHV8 members on one hand, and poxviral VacV and OrfV members on the other. These similarities among closely related viruses might also suggest that the interaction network of each TMD is important for the survival of the virus so that it has been conserved throughout its evolution. Otherwise, the interaction pattern of each TMD would have been lost through the fast mutation rate and high evolutionary pressure scenario associated with the virus. Sequence similarity would have been expected to accompany similarities in the vBcl2 TMD interaction pattern. However, as the sequence alignment exposed, in the TMDs of the analyzed proteins, no sequence homology pattern could explain the observed results. This finding implies the presence of a structural pattern underlying the sequence that could govern the TMD–TMD interactions and that has been conserved throughout viral evolution. Of note, the similarities in the interaction network between closely related viruses also suggest that the observed interactions are not the result of random contacts among TMDs resulting from non-specific affinities or overcrowding.

Our assays allowed us to analyze each potential interaction independently. However, for each vBcl2 protein in vivo, all the TMD–TMD interactions with cBcl2s could occur at once. It remains to be confirmed whether there is a preferred (hierarchical) interaction that is responsible for the anti-apoptotic effect or there are several TMD–TMD interactions that occur side-by-side, all of them contributing to the cell death blockage. Currently, data exist to support either possibility. On the one hand, vBcl2s can interact with multiple cBcl2s. For example, as recently demonstrated, HHV4 drives chemoresistance and lymphomagenesis by inhibiting multiple cellular pro-apoptotic proteins[61]. On the other hand, a hierarchic interaction profile for the TMDs of cBcl2 has also been observed[63]. These two possibilities are not exclusive and most likely occur simultaneously in the cells.

According to our results, vBcl2 TMD are capable of forming homo- and hetero-oligomers. These two types of interactions might not be exclusive but collaborative/synergistic. The techniques we used for the study of hetero-oligomers (BiFC and BlaTM) do not impede the formation of homo-oligomers alongside them. In these assays, the formation of a heterodimer is required to obtain fluorescence (BiFC) or antibiotic resistance (BlaTM). Nonetheless, it is possible (even probable) that homo-oligomers are formed prior to the interaction between a viral and a cellular TMD. Indeed, VacV has been shown to adopt a dimeric fold through its soluble region which enables hetero-oligomeric binding to pro-apoptotic members of the Bcl2 family[64]. In this case, it is likely that the TMD first participates in the formation of the homodimer and then in the interaction with cBcl2.

Next, we focused on the TMD–TMD interaction between the strong partners Bcl2-HHV8 and Bax-MyxV. Bcl2 and HHV8 TMDs were detected in mitochondrial membranes where Bcl2 is thought to interact with pro-apoptotic members of the Bcl2 family to control apoptosis. Whether HHV8 requires its interaction with Bcl2 to exert its anti-apoptotic function remains to be seen. Our data also revealed that the presence of the soluble domains of Bcl2 did not affect TMD–TMD interaction between Bcl2 and HHV8 (Fig. 4a), reinforcing the idea that intramembrane interactions could occur side-by-side with contacts through soluble domains. With respect to the MyxV-Bax intramembrane interaction, we not only confirmed the interaction seen with the BiFC assay but also demonstrated that MyxV TMD can inhibit Bax TMD-induced apoptosis (Fig. 5). Furthermore, MyxV-FL but not MyxV-T20 could rescue Bax-FL-induced apoptosis. These results might imply that MyxV can exert its anti-apoptotic action through direct interactions with pro-apoptotic Bax (in agreement with the previous results[43]) and that its TMD is a key domain that plays an active role in the control of cell death. Nonetheless, other scenarios in which a third protein is involved cannot be ruled out. Of interest, the in silico analysis of both the putative HHV8-Bcl2 and the MyxV-Bax intramembrane dimers revealed ridge–groove arrangements between small and large residues in opposing TMDs. This type of interaction has been observed previously in other TMD dimers, particularly in the case of the GpA homodimer.

Finally, our results, using both drug and viral apoptotic stimuli, indicated that the TMD in the vBcl2s is necessary for the control of apoptosis. Furthermore, when HHV8, MyxV, or VacV lose their ability to interact with cellular proteins through the TMD, they become incapable of regulating apoptosis, as demonstrated when the TMD in the vBcl2 was deleted or substituted by the TMD of T20 (Figs. 7 and 8). These results open the way to the design of new antivirals that could interfere with these hosts–viral TMD–TMD interactions. Of interest, MyxV-FL behaved as a strong apoptosis inhibitor regardless of the stimuli. In contrast, HHV8-FL worked better as an apoptosis inhibitor when PKR (under a VacV promoter) was used as the stimulus. This divergence between MyxV and HHV8 suggests a difference in their mechanism of action, which could also be inferred by analyzing the interaction profile of their TMDs (Fig. 6 and Supplementary Fig. 10).

In conclusion, we have identified the Ct hydrophobic region of the vBcl2 as a true TMD that can interact with cBcl2 TMDs. We also have demonstrated that these intramembrane interactions are crucial for the viral control of cell fate. This work advances our understanding of how viruses control cellular apoptosis for their advantage and how apoptosis is regulated in the cell. Furthermore, the interactions described here expand knowledge about how viruses interact with their host.

Methods

Cell cultures, plasmids, and reagents. Human embryonic kidney 293T cells (HEK 293T), human colorectal carcinoma-derived cells (HCT 116), and human epithelial cervical cancer cells (HeLa) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), and penicillin–streptomycin (P/S) (100 U/mL) (Gibco). All cells were grown at 37 °C, 5% CO_2.

The TMD sequences were synthesized by Invitrogen (GeneArt gene synthesis), PCR amplified, and subcloned into the appropriated vector either using a standard digestion-ligation protocol or using the InFusion cloning system following the manufacturer's protocol (Takara). Mutations into the TMD were introduced by site-directed mutagenesis using the Quick Change II kit following the manufacturer's instructions (Agilent Technologies). A full list of all the primers can be found in Supplementary Table 1. All DNA manipulations were confirmed by the sequencing of plasmid DNAs (Macrogen). Transfection of DNA into eukaryotic cells was performed in Opti-MEM reduced serum medium (Gibco) with Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications. The recommendations of the International Committee on Taxonomy of Viruses (http://www.ictvonline.org/index.asp) were used as guidelines for the viral nomenclature.

In vitro transcription and translation. The Lep-derived constructs were assayed using the TNT T7 Quick Coupled System (#L1170, Promega). Each reaction containing 1 µL of PCR product, 0.5 µL of EasyTag™ EXPRESS 35S Protein Labeling Mix (Perkin Elmer) (5.5 µCi), and 0.3 µL of microsomes (tRNA Probes) was incubated at 30 °C for 90 min. Samples were analyzed by SDS-PAGE. The bands were quantified using a Fuji FLA-3000 phosphoimager and the Image Reader 8.1 software. Free energy was calculated using: $AG_{app} = -RT$ lnK_{app}, where $K_{app} = f2g/f1g$ being f1g and f2g the fraction of single glycosylated and double glycosylated protein, respectively. Endoglycosidase H treatment (Roche) was carried according to the specifications of the manufacturer.

Bimolecular fluorescent complementation (BiFC) assay. For the generation of BiFC chimeric plasmids including the Nt or Ct of the Venus Fluorescent Protein (VN, VC, respectively) Addgene #27097, #22011 (a gift from Chang-Deng H)[65] plasmids were modified to clone the cellular and viral Bcl2 TMDs at the Ct of the VFP. Chimeras (500 ng VN + 500 ng VC) were transfected into 2×10^5 HEK 293T cells together with a plasmid expressing *Renilla luciferase* under the CMV promoter (pRL-CMV) (50 ng) for signal normalization. Additionally, for the competition assay, 500 ng of a plasmid encoding Bcl2 full-length or Bcl2 lacking the TMD were transfected. Cells were incubated at 37 °C, 5% CO_2 for 48 h, PBS washed and collected for fluorescence and luciferase measurements (Victor X3 plate reader). For the Renilla luciferase readings, we used the Renilla Luciferase Glow Assay Kit (Pierce, Thermofisher) according to the manufacturer's protocol. In each experiment, the fluorescence/luminescence ratio obtained with the GpA homodimer was used as a 100% oligomerization value and the rest of the values adjusted accordingly. All experiments were done at least in triplicates.

BlaTM assay. Competent *E. coli* BL21-DE3 cells were co-transformed with N-BLa and C-BLa plasmids, version 1.1[39], containing a given TMD pair and grown overnight at 37 °C on LB-agar plates containing 34 µg/mL of chloramphenicol (Cm) and 35 µg/mL of kanamycin (Kan) for plasmid inheritance. After o/n incubation at 37 °C, colonies were picked for immediate use or the plates were sealed with Parafilm (Pechiney Plastic Packaging) and stored at 4 °C for up to one week. Overnight cultures were conducted by inoculating 5 mL of LB-medium (Cm, Kan) with 10 colonies from one agar plate, followed by o/n incubation in an orbital incubator at 37 °C, 200 rpm. An expression culture was started with a 1:10 dilution of the overnight culture in 4 mL expression medium: LB-medium (Cm, Kan) containing 1.33 mM arabinose. After 4 h at 37 °C, the expression cultures were diluted to an $OD_{600} = 0.1$ in expression medium. To expose the bacteria to different ampicillin concentrations, an LD_{50} culture was prepared by pipetting 100 µL of the diluted expression culture into each cavity of a 96-deep well plate (96 square well, 2 mL, VWR) containing 400 µL of expression media (final OD_{600} = 0.02). Freshly prepared ampicillin stock (100 mg/mL in ethanol) was added, resulting in ampicillin concentrations ranging from 0 to 350 µg/mL, depending on the affinity of the TMD under investigation. As a rule, the maximum ampicillin concentration to be used for a particular case should be about twice the mean LD_{50}. The plates were incubated in a moisturized container for 16 h at 37 °C and 250 rpm on a shaker (shaking amplitude 10 mm, KS 260 Basic, IKA) containing tips in every well to ensure a proper agitation. Cell density was measured via absorbance at 544 nm in a microplate reader (Victor X3, Perkin Elmer). To minimize clonal variation, at least two transformations were done and at least two separate LD_{50} cultures were inoculated from each batch of transformed bacteria using ten colonies for each culture. Thus, at least 40 colonies entered each determination of LD_{50}. To measure and collect LD_{50} values from the dose–response curves, we used Prism 8 from GraphPad.

To analyze the expression levels of the chimeras, competent *E. coli* BL21-DE3 cells were transformed with one N-BLa or C-BLa plasmid, version 1.1, containing a given TMD and grown overnight at 37 °C on LB-agar plates containing 34 µg/mL

of Cm (for N-BLa) or 35 µg/mL of Kan (for C-BLa) for plasmid inheritance. After o/n incubation at 37 °C, cultures were conducted by inoculating 5 mL of LB-medium (Cm or Kan) with 10 colonies from one agar plate, followed by o/n incubation in an orbital incubator at 37 °C, 200 rpm. An expression culture was started with a 1:10 dilution from the overnight culture in 4 mL expression medium: LB-medium (Cm or Kan) containing 1.33 mM arabinose. After 4 h at 37 °C, the expression cultures were diluted to an $OD_{600} = 0.1$ in 5 ml of expression medium (final volume) and grown o/n, 37 °C, 200xg. The morning after, 100 µL from each culture were transferred to a black 96-well plate to measure the fluorescence (Victor X3, Perkin Elmer).

Cell-viability assays. To measure doxorubicin-induced apoptosis 1.5×10^6 HeLa cells were plated in a 24 wells plate containing 0.5 ml of media in each well. After overnight incubation, each well was transfected in triplicates with 500 ng of DNA. After two days of expression, cells were treated with doxorubicin (stock 2 mM in DMSO) achieving a final concentration of 15 µM. Approximately, 16 h post-treatment cells (including those in the supernatant) were collected and their viability was measured using Trypan blue and using an automated cell counter (Invitrogen, Countess™ II). At least, 2 measurements per well expression done. Additionally, the viability and apoptotic state of the cells was measured using the BD Annexin V: FITC Apoptosis Detection Kit II according to the manufacturer specifications and a BD LSR-Fortessa (Beckton Dickinson). All flow cytometry countings were done at the tissue culture and flow cytometry core facilities of the SCSIE (University of Valencia) following the recommendations of the BD Annexin V: FITC Apoptosis Detection Kit II. To measure Bax TMD-induced apoptosis HCT 116 cells were transfected with 250 ng of the plasmid encoding VN-Bax and 250 ng of either VC-H2, VC-Bcl2 or VC-MyxV encoding plasmids or mock-transfected. After 48 h, cells were trypsinized and viability was measured twice per well using Trypan Blue and automated Countess™ II cell counter (Invitrogen). The percentage of survival was calculated using the mock-transfected cells as the reference value.

To measure Bax-induced apoptosis, 3×10^6 HeLa cells were plated in a 24-well plate (0.5 ml of media in each well). After overnight incubation, cells were transfected (in triplicate) with 500 ng of either MyxV-FL, M11L-T20, Bcl2-FL, Bcl2-T20, or empty pCAGGS (negative control). Additionally, cells were co-transfected with 400 ng of Bax DNA. For the positive control, 500 ng of Bcl2-FL and 400 ng of empty pCAGGS were transfected. After 24 hrs of expression, cells were collected and their viability was measured using Trypan blue and an automated cell counter (Invitrogen, Countess™ II). At least 2 measurements were done per well.

Bio-informatic resources. Prediction of the TMD on full-length sequences was done either with the ΔG prediction server v1.0[31,32] or the TMHMM server v2.0[29,30] using default parameters. Principal Component Analysis (PCA) was done with the Gene Expression Similarity Investigation Suite (Genesis)[66] (http://genome.tugraz.at) using the default parameters. For the multiple sequence alignment, *Clustal Omega* was launched from the EMBL-EBI site[67]. Contact surface predictions for TM dimers were done with PREDDIMER[67] (http://model.nmr.ru/preddimer/). From all the options provided by the algorithm, only the lowest energy option was considered.

Confocal microscopy. Confocal micrographs were done at the Microscopy Core Facility of the SCSIE (University of Valencia) using an Olympus FV1000 confocal microscope with a ×60 oil lens. ER (Sec61β), plasma membrane (the first twenty residues of neuromodulin) and, mitochondria (Tomm20) mCherry fluorescent-labeled markers were obtained from Addgene plasmid repository #49155[68], #55779[69] and #55146 (a gift from Michael Davidson, Institute of Molecular Biophysics and Center for Materials Research and Technology, The Florida State University) respectively. Cells (5×10^3 cells/well) were seeded on 10 mm coverslides treated with poly-Lys and placed in 24-well plates. The next day, cells were transfected with the appropriate plasmids. After 24 h, the cells were fixed (4% paraformaldehyde) and DAPI stained before image capture. A 1:1000 dilution in TBS 0.005% Tween Rabbit anti c-Myc (Sigma PLA0001) antibody followed by an anti-Rabbit Alexa 488 conjugated (Life Technologies A21206) (1:1000) was used to label viral proteins. Pictures were taken in an Olympus FV1000 confocal microscope. Laser intensity was individually adjusted in all samples. Pictures were not used for quantification.

Viruses and infections. The VacV recombinants used in this study have been previously described: VacV-PKR (previously named as VV-p68)[48], which expresses the dsRNA-dependent protein kinase upon induction of the isopropyl β-D-thiogalactosidase (IPTG)-dependent inducible promoter, VT7LacOI that expresses the T7 RNA polymerase in an IPTG-dependent manner, the parental MVA-C and the deletion mutant MVA-C-ΔF1L lacking the viral F1L gene[49]. To express viral and cBcl2 proteins alongside viral infection, the selected DNA sequences were cloned into pVOTE.2[70] plasmid under the control of the IPTG-dependent inducible T7 promoter. Infections were performed on pre-confluent (<75%) cell monolayers. Cultures were co-infected with VT7LacOI and either VV-PKR, MVA-C, or MVA-C-ΔFIL diluted in DMEM to a multiplicity of infection (MOI) of 2.5 PFU/cell of each virus. After the one-hour adsorption cells were transfected with 50 ng of the

indicated plasmid. Transfections of infected cells were performed with Lipofecta-mine 2000 reagent (Invitrogen), following the protocol recommended by the supplier. Infected cells were incubated in the presence of IPTG (5 mM) at 37 °C until 24 h.p.i.

Caspase 3/7 activity assay. Quantification of caspase activity was carried out by using the Caspase-Glo 3/7 assay kit (Promega) following the protocol recommended by the supplier. HeLa cell monolayers grown in 24-well plates were infected and transfected, in duplicate, under the conditions indicated above. At 24 h.p.i. cells were harvested in medium and kept frozen until their analysis. 25 µL of the cell lysates under study was added to 25 µL of Caspase-Glo 3/7 reagent in a 96-well plate. Plates were gently shaken and then incubated in the dark at room temperature for 60 min before recording the luciferase activity by using an Appliskan luminometer (ThermoScientific).

Western blot analysis. Cell monolayers were lysed in Laemmli´s sample buffer (62.5 mM Tris-HCl [pH 6.8], 2% sodium dodecyl sulfate [SDS], 0.01% bromo-phenol blue, 10% glycerol and 5% β-mercaptoethanol). Protein samples were subjected to 12% SDS-polyacrylamide gel electrophoresis (PAGE) and transferred to nitrocellulose membranes (BioRad). Membranes were blocked for 30 min at room temperature in Tris-buffered saline supplemented with 0.05% Tween 20 (TBS-T) containing 5% non-fat dry milk, and later incubated with primary anti-bodies diluted in the same buffer at 4 °C overnight. Antibodies used in this study were β-actin (Santa Cruz Biotechnology SC-47778), c-Myc (Sigma PLA0001 or Roche 11667149001), Histone 3 (Sigma H0164), and Flag (Sigma B3111). Then, membranes were washed with TBS-T and incubated with goat anti-mouse IgG-peroxidase conjugate (Sigma DC02L) for 1 h at room temperature and washed again. All antibodies were used at a 1:10,000 dilution in TBS-T with 5% non-fat dry milk. Detection of immunoreactive proteins was carried out using the enhanced chemiluminescence (ECL) reaction (SuperSignal ThermoScientific) and detected by the ChemiDoc Touch Imaging System (BioRad).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

References
1. Kerr, J. F., Wyllie, A. H. & Currie, A. R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *Br. J. Cancer* **26**, 239–257 (1972).
2. Hua, F., Li, K., Shang, S., Wang, F. & Hu, Z. Immune signaling and autophagy regulation. *Adv. Exp. Med. Biol.* **1206**, 551–593 (2019).
3. Menzies, F. M. et al. Autophagy and neurodegeneration: pathogenic mechanisms and therapeutic opportunities. *Neuron* **93**, 1015–1034 (2017).
4. Häcker, G. Apoptosis in infection. *Microbes Infect.* **20**, 552–559 (2018).
5. Galluzzi, L., Brenner, C., Morselli, E., Touat, Z. & Kroemer, G. Viral control of mitochondrial apoptosis. *PLoS Pathog.* **4**, e1000018 (2008).
6. Ucker, D. S. Exploiting death: apoptotic immunity in microbial pathogenesis. *Cell Death Differ.* **23**, 990–996 (2016).
7. Kim, H. et al. Hierarchical regulation of mitochondrion-dependent apoptosis by BCL-2 subfamilies. *Nat. Cell Biol.* **8**, 1348–1358 (2006).
8. Boise, L. H. et al. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**, 597–608 (1993).
9. Oltvai, Z. N., Milliman, C. L. & Korsmeyer, S. J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell* **74**, 609–619 (1993).
10. Wang, K., Yin, X. M., Chao, D. T., Milliman, C. L. & Korsmeyer, S. J. BID: a novel BH3 domain-only death agonist. *Genes Dev.* **10**, 2859–2869 (1996).
11. Andreu-Fernández, V. et al. The C-terminal domains of apoptotic BH3-only proteins mediate their insertion into distinct biological membranes. *J. Biol. Chem.* **291**, 25207–25216 (2016).
12. Delbridge, A. R. D., Grabow, S., Strasser, A. & Vaux, D. L. Thirty years of BCL-2: translating cell death discoveries into novel cancer therapies. *Nat. Rev. Cancer* **16**, 99–109 (2016).
13. Kelekar, A., Chang, B. S., Harlan, J. E., Fesik, S. W. & Thompson, C. B. Bad is a BH3 domain-containing protein that forms an inactivating dimer with Bcl-XL. *Mol. Cell. Biol.* **17**, 7040–7046 (1997).
14. Xie, Z., Schendel, S., Matsuyama, S. & Reed, J. C. Acidic pH promotes dimerization of Bcl-2 family proteins. *Biochemistry* **37**, 6410–6418 (1998).
15. O'Neill, J. W., Manion, M. K., Maguire, B. & Hockenbery, D. M. BCL-XL dimerization by three-dimensional domain swapping. *J. Mol. Biol.* **356**, 367–381 (2006).
16. Cosentino, K. & García-Sáez, A. J. Bax and Bak pores: are we closing the circle? *Trends Cell Biol.* **27**, 266–275 (2017).
17. Andreu-Fernández, V. et al. Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. *Proc. Natl Acad. Sci. USA* https://doi.org/10.1073/pnas.1612322114 (2016).
18. Polčic, P., Mentel, M., Gavurníková, G. & Bhatia-Kiššová, I. To keep the host alive - the role of viral Bcl-2 proteins. *Acta Virol.* **61**, 240–251 (2017).
19. Kvansakul, M., Caria, S. & Hinds, M. G. The Bcl-2 family in host-virus interactions. *Viruses* **9**, 209 (2017).
20. Neumann, S. et al. How do viruses control mitochondria-mediated apoptosis? *Virus Res.* **209**, 45–55 (2015).
21. Kvansakul, M. & Hinds, M. G. Structural biology of the Bcl-2 family and its mimicry by viral proteins. *Cell Death Dis.* **4**, e909 (2013).
22. Pearson, G. R. et al. Identification of an Epstein-Barr virus early gene encoding a second component of the restricted early antigen complex. *Virology* **160**, 151–161 (1987).
23. Cheng, E. H. et al. A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. *Proc. Natl Acad. Sci. USA* **94**, 690–694 (1997).
24. Bellows, D. S. et al. Antiapoptotic herpesvirus Bcl-2 homologs escape caspase-mediated conversion to proapoptotic proteins. *J. Virol.* **74**, 5024–5031 (2000).
25. Wasilenko, S. T., Stewart, T. L., Meyers, A. F. A. & Barry, M. Vaccinia virus encodes a previously uncharacterized mitochondrial-associated inhibitor of apoptosis. *Proc. Natl Acad. Sci. USA* **100**, 14345–14350 (2003).
26. Nichols, D. B., De Martini, W. & Cottrell, J. Poxviruses utilize multiple strategies to inhibit apoptosis. *Viruses* **9**, 215 (2017).
27. Douglas, A. E., Corbett, K. D., Berger, J. M., McFadden, G. & Handel, T. M. Structure of M11L: a myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2. *Protein Sci. Publ. Protein Soc.* **16**, 695–703 (2007).
28. Westphal, D. et al. A novel Bcl-2-like inhibitor of apoptosis is encoded by the parapoxvirus ORF virus. *J. Virol.* **81**, 7178–7188 (2007).
29. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
30. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182 (1998).
31. Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature* **450**, 1026–1030 (2007).
32. Hessa, T. et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
33. Braunger, K. et al. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. *Science* **360**, 215–219 (2018).
34. Kerppola, T. K. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat. Protoc.* **1**, 1278–1286 (2006).
35. Wouters, E., Vasudevan, L., Crans, R. A. J., Saini, D. K. & Stove, C. P. Luminescence- and fluorescence-based complementation assays to screen for GPCR oligomerization: current state of the art. *Int. J. Mol. Sci.* **20**, 2958 (2019).
36. MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. A transmembrane helix dimer: structure and implications. *Science* **276**, 131–133 (1997).
37. Orzáez, M., Pérez-Payá, E. & Mingarro, I. Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. *Protein Sci. Publ. Protein Soc.* **9**, 1246–1253 (2000).
38. Lemmon, M. A. et al. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J. Biol. Chem.* **267**, 7683–7689 (1992).
39. Schanzenbach, C., Schmidt, F. C., Breckner, P., Teese, M. G. & Langosch, D. Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions. *Sci. Rep.* **7**, 43476 (2017).
40. Popgeorgiev, N., Jabbour, L. & Gillet, G. Subcellular localization and dynamics of the Bcl-2 family of proteins. *Front. Cell Dev. Biol.* **6**, 13 (2018).
41. Polyansky, A. A. et al. PREDDIMER: a web server for prediction of transmembrane helical dimers. *Bioinformatics* **30**, 889–890 (2014).
42. Sulistijo, E. S. & MacKenzie, K. R. Sequence dependence of BNIP3 transmembrane domain dimerization implicates side-chain hydrogen bonding and a tandem GxxxG motif in specific helix-helix interactions. *J. Mol. Biol.* **364**, 974–990 (2006).

43. Su, J. et al. Myxoma virus M11L blocks apoptosis through inhibition of conformational activation of Bax at the mitochondria. *J. Virol.* **80**, 1140–1151 (2006).
44. Wang, G. et al. Myxoma virus M11L prevents apoptosis through constitutive interaction with Bak. *J. Virol.* **78**, 7097–7111 (2004).
45. Rooswinkel, R. W. et al. Antiapoptotic potency of Bcl-2 proteins primarily relies on their stability, not binding selectivity. *Blood* **123**, 2806–2815 (2014).
46. Kvansakul, M. et al. A structural viral mimic of prosurvival Bcl-2: a pivotal role for sequestering proapoptotic Bax and Bak. *Mol. Cell* **25**, 933–942 (2007).
47. Lee, S. B. & Esteban, M. The Interferon-induced double-stranded RNA-activated protein kinase induces apoptosis. *Virology* **199**, 491–496 (1994).
48. Lee, S. B., Rodríguez, D., Rodríguez, J. R. & Esteban, M. The apoptosis pathway triggered by the interferon-induced protein kinase PKR requires the third basic domain, initiates upstream of Bcl-2, and involves ICE-like proteases. *Virology* **231**, 81–88 (1997).
49. Perdiguero, B. et al. Deletion of the viral anti-apoptotic gene F1L in the HIV/AIDS vaccine candidate MVA-C enhances immune responses against HIV-1 antigens. *PLoS ONE* **7**, e48524 (2012).
50. Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L. & Kroemer, G. Immunogenic cell death in cancer and infectious disease. *Nat. Rev. Immunol.* **17**, 97–111 (2017).
51. Müller-Coan, B. G., Caetano, B. F. R., Pagano, J. S. & Elgui de Oliveira, D. Cancer progression goes viral: the role of oncoviruses in aggressiveness of malignancies. *Trends Cancer* **4**, 485–498 (2018).
52. Plummer, M. et al. Global burden of cancers attributable to infections in 2012: a synthetic analysis. *Lancet Glob. Health* **4**, e609–616 (2016).
53. Ferrara, P., Agüero, F., Masuet-Aumatell, C. & Ramon-Torrell, J. M. Burden of cancer mortality attributable to carcinogenic infections in Spain. *Med. Clin.* https://doi.org/10.1016/j.medcli.2019.11.005 (2020).
54. Letai, A. Apoptosis and cancer. *Annu. Rev. Cancer Biol.* **1**, 275–294 (2017).
55. Grau, B. et al. The role of hydrophobic matching on transmembrane helix packing in cells. *Cell Stress* **1**, 90–106 (2017).
56. Anbazhagan, V. & Schneider, D. The membrane environment modulates self-association of the human GpA TM domain–implications for membrane protein folding and transmembrane signaling. *Biochim. Biophys. Acta* **1798**, 1899–1907 (2010).
57. Bocharov, E. V. et al. Helix-helix interactions in membrane domains of bitopic proteins: specificity and role of lipid environment. *Biochim. Biophys. Acta Biomembr.* **1859**, 561–576 (2017).
58. Peña-Blanco, A. & García-Sáez, A. J. Bax, Bak and beyond—mitochondrial performance in apoptosis. *FEBS J.* **285**, 416–431 (2018).
59. Kalbfleisch, T., Cambon, A. & Wattenberg, B. W. A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic Cph. Den.* **8**, 1687–1694 (2007).
60. Martinez-Gil, L. & Mingarro, I. Viroporins, examples of the two-stage membrane protein folding model. *Viruses* **7**, 3462–3482 (2015).
61. Fitzsimmons, L. et al. EBV BCL-2 homologue BHRF1 drives chemoresistance and lymphomagenesis by inhibiting multiple cellular pro-apoptotic proteins. *Cell Death Differ.* https://doi.org/10.1038/s41418-019-0435-1 (2019).
62. Postigo, A., Cross, J. R., Downward, J. & Way, M. Interaction of F1L with the BH3 domain of Bak is responsible for inhibiting vaccinia-induced apoptosis. *Cell Death Differ.* **13**, 1651–1662 (2006).
63. Bleicken, S., Hantusch, A., Das, K. K., Frickey, T. & Garcia-Saez, A. J. Quantitative interactome of a membrane Bcl-2 network identifies a hierarchy of complexes for apoptosis regulation. *Nat. Commun.* **8**, 73 (2017).
64. Campbell, S. et al. Structural insight into BH3 domain binding of vaccinia virus antiapoptotic F1L. *J. Virol.* **88**, 8667–8677 (2014).
65. Kodama, Y. & Hu, C.-D. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. *BioTechniques* **49**, 793–805 (2010).
66. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics Oxf. Engl.* **18**, 207–208 (2002).
67. F, S. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2010).
68. Zurek, N., Sparks, L. & Voeltz, G. Reticulon short hairpin transmembrane domains are used to shape ER tubules. *Traffic Cph. Den.* **12**, 28–41 (2011).
69. Yost, E. A., Mervine, S. M., Sabo, J. L., Hynes, T. R. & Berlot, C. H. Live cell analysis of G protein beta5 complex formation, function, and targeting. *Mol. Pharmacol.* **72**, 812–825 (2007).
70. Ward, G. A., Stover, C. K., Moss, B. & Fuerst, T. R. Stringent chemical and thermal regulation of recombinant gene expression by vaccinia virus vectors in mammalian cells. *Proc. Natl Acad. Sci. USA* **92**, 6773–6777 (1995).

Author contributions

L.M.-G., I.M., and D.R. designed the experiments. M.J.G.-M., G.D., B.G., E.D.-B., and L.M.-G. performed the experiments. All authors contributed to the analysis of the data. L.M.-G. wrote and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-19881-9.

Correspondence and requests for materials should be addressed to L.M.-G.

Peer review information *Nature Communications* thanks Georg Häcker, Marc Kvansakul and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

Supplementary Files

Maria Jesús García-Murria[1,‡], Gerard Duart[1,‡], Brayan Grau[1], Elisabet Diaz-Beneitez[2], Dolores Rodríguez[2], Ismael Mingarro[1], Luis Martínez-Gil[1*]

[1]Department of Biochemistry and Molecular Biology, Institut de Biotecnologia i Biomedicina, Universitat de València, 46100 Burjassot, Spain.

[2]Department of Molecular and Cell Biology, Centro Nacional de Biotecnología Consejo Superior de Investigaciones Científicas, Campus Universidad Autónoma, 28049 Madrid, Spain.

[‡]These authors contributed equally

[*]Author to whom correspondence might be addressed. Luis Martínez-Gil, Luis.martinez-gil@uv.es

```
Herpesvirus
```

>G3CKQ0 **HHV4** Apoptosis regulator BHRF1
MAYSTREILLALCIRDSRVHGNGTLHPVLELAARETPLRLSPEDTVVLRYHVLLEEIIERNSETFTETWNRFITHTEHLD
LDFNSVFLEIFHRGDPSLGRALAWMAWCMHACRTLCCNQSTPYYVVDLSVRGMLEASEGLDGWIHQQGGWSTLIEDNIPG
SRR**FSWTLFLAGLTLSLLVICSYLFI**SRGRH
Predicted ΔG -2.632

>F5HGJ3 **HHV8** Apoptosis regulator Bcl-2 homolog
MDEDVLPGEVLAIEGIFMACGLNEPEYLYHPLLSPIKLYITGLMRDKESLFEAMLANVRFHSTTGIQLGLSMLQVSGDGN
MNWGRALAILTFGSFVAQKLSNEPHLRDFALAVLPVYAYEAIGPQWFRARGGWRGLKAYCTQVLTRRRG**RRMTALLGSIA**
LLATILAAVAMSRR
Predicted ΔG -1.008

>Q9WH78 **BoHV4** V-Bcl-2-like protein
MSLFFVVWYWVNYITKVCSGEVYIPSVLKFQYHSDTEHEPYSNLCKNLITMAEQDMDEVVSTIRRLLVECGMGLEEYLEH
PVTAPIKVAVQDVIRTKQDIFSNFLTNINSVEDLETLGHAITTLNDYPSPNMGRVVCGIAFSVYVVQTVCKRKPLLVRCC
LDIFTRATVQALNVNWFLQEGGWPALASFCKVVNSPSPRS**RWLFPMFAISGLVLTVGVAR**NMVHFT
Predicted ΔG -0.031

```
Poxvirus
```

>Q77PA8 **MyxV** Apoptosis regulator M11L
MMSRLKTAVYDYLNDVDITECTEMDLLCQLSNCCDFINETYAKNYDTLYDIMERDILSYNIVNIKNTLTFALRDASPSVK
LATLTLLASVIKKLNKIQHTDAAMFSEVIDGIVAEEQQVIGFIQKKCKYNTTYYNVRS<u>GGCK</u>**ISVYLTAAVVGFVAYGIL**
KWYRGT
Predicted ΔG -0.217

>P24356 **VacV** Protein F1L
MLSMFMCNNIVDYVDDIDNGIVQDIEDEASNNVDHDYVYPLPENMVYRFDKSTNILDYLSTERDHVMMAVRYYMSKQRLD
DLYRQLPTKTRSYIDIINIYCDKVSNDYNRDMNIMYDMASTKSFTVYDINNEVNTILMDNKGLGVRLATISFITELGRRC
MNPVETIKMFTLLSHTICDDYFVDYITDISPPDNTIPNTST**REYLKLIGITAIMFATYKTL**KYMIG
Predicted ΔG 0.902

>A0A0R8HV90 **OrfV** Apoptosis inhibitor
MANRDDIDASAVMAAYLAREYAEAVEEQLTPRERDALEALRVSGEEVRSPLLQELSNAGEHRANPENSHIPAALVSALLE
APTSPGRMVTAVELCAQMGRLWTRGRQLVDFMRLVYVLLDRLPPTADEDLGAWLQAVARVHGTRRRL<u>H</u>**RALGVGAVVAGV**
GMLLLGVRVL*R*RT
Predicted ΔG -0.815

Supplementary Figure 1. **vBcl2 sequences**. The figure includes the sequences (in a FASTA format) of the HHV4 BHRF1, HHV8 ORF16, BoHV4 ORF16, MyxV M11L, VacV F1L, and OrfV ORFV125 proteins. The Uniprot code is shown alongside the viral acronym and the protein name. The TMDs identified by the ΔG prediction server using the full protein scan mode are highlighted in yellow, the TMDs predicted by the TMHMM Server are underlined, and the hydrophobic sequences used in our assays are in bold. Included below each sequence is the predicted ΔG, calculated using the ΔG prediction server.

```
TMD Incorporated in our assays

>P10415 Bcl2 Apoptosis regulator Bcl-2
MAHAGRTGYDNREIVMKYIHYKLSQRGYEWDAGDVGAAPPGAAPAPGIFSSQPGHTPHPAASRDPVARTSPLQTPAAPGA
AAGPALSPVPPVVHLTLRQAGDDFSRRYRRDFAEMSSQLHLTPFTARGRFATVVEELFRDGVNWGRIVAFFEFGGVMCVE
SVNREMSPLVDNIALWMTEYLNRHLHTWIQDNGGWDAFVELYGPSMRPLFDFSWLSLKTLLSLALVGACITLGAYLGHK

>Q07817 BclXL Bcl-2-like protein 1
MSQSNRELVVDFLSYKLSQKGYSWSQFSDVEENRTEAPEGTESEMETPSAINGNPSWHLADSPAVNGATGHSSSLDAREV
IPMAAVKQALREAGDEFELRYRRAFSDLTSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQ
VLVSRIAAWMATYLNDHLEPWIQENGGWDTFVELYGNNAAAESRKGQERFNRWFLTGMTVAGVVLLGSLFSRK

>Q07812 Bax Apoptosis regulator BAX
MDGSGEQPRGGGPTSSEQIMKTGALLLQGFIQDRAGRMGGEAPELALDPVPQDASTKKLSECLKRIGDELDSNMELQRMI
AAVDTDSPREVFFRVAADMFSDGNFNWGRVVALFYFASKLVLKALCTKVPELIRTIMGWTLDFLRERLLGWIQDQGGWDG
LLSYFGTPTWQTVTIFVAGVLTASLTIWKKMG

>Q16611 Bak Bcl-2 homologous antagonist/killer
MASGQGPGPPRQECGEPALPSASEEQVAQDTEEVFRSYVFYRHQQEQEAEGVAAPADPEMVTLPLQPSSTMGQVGRQLAI
IGDDINRRYDSEFQTMLQHLQPTAENAYEYFTKIATSLFESGINWGRVVALLGFGYRLALHVYQHGLTGFLGQVTRFVVD
FMLHHCIARWIAQRGGWVAALNLGNGPILNVLVVLGVVLLGQFVVRRFFKS

>Q96LC9 Bmf Bcl-2-modifying factor
MEPSQCVEELEDDVFQPEDGEPVTQPGSLLSADLFAQSLLDCPLSRLQLFPLTHCCGPGLRPTSQEDKATQTLSPASPSQ
GVMLPCGVTEEPQRLFYGNAGYRLPLPASFPAVLPIGEQPPEGQWQHQAEVQIARKLQCIADQFHRLHVQQHQQNQNRVW
WQILLFLHNLALNGEENRNGAGPR

>Q13323 Bik Bcl-2-interacting killer
MSEVRPLSRDILMETLLYEQLLEPPTMEVLGMTDSEEDLDPMEDFDSLECMEGSDALALRLACIGDEMDVSLRAPRLAQL
SEVAMHSLGLAFIYDQTEDIRDVLRSFMDGFTTLKENIMRFWRSPNPGSWVSCEQVLLALLLLLALLLPLLSGGLHLLLK

>Q13794 Noxa Phorbol-12-myristate-13-acetate-induced protein 1
MPGKKARKNAQPSPARAPAELEVECATQLRRFGDKLNFRQKLLNLISKLFCSGT

>Q15388 Tomm20 Mitochondrial import receptor subunit TOM20 (Human)
MVGRNSAIAAGVCGALFIGYCIYFDRKRRSDPNFKNRLRERRKKQKLAKERAGLSKLPDLKDAEAVQKFFLEEIQLGEEL
LAQGEYEKGVDHLTNAIAVCGQPQQLLQVLQQTLPPPVFQMLLTKLPTISQRIVSAQSLAEDDVE
Predicted ΔG 0.217

>Q9NS69 Tomm22 Mitochondrial import receptor subunit TOM22 (Human)
MAAAVAAAGAGEPQSPDELLPKGDAEKPEEELEEDDDEELDETLSERLWGLTEMFPERVRSAAGATFDLSLFVAQKMYRF
SRAALWIGTTSFMILVLPVVFETEKLQMEQQQQLQQRQILLGPNTGLSGGMPGALPSLPGKI

>P00803 Lep Leader Signal peptidase I (E.coli)
MANMFALILVIATLVTGILWCVDKFFFAPKRRERQAAAQAAAGDSLDKATLKKVAPKPGWLETGASVFPVLAIVLIVRSF
IYEPFQIPSGSMMPTLLIGDFILVEKFAYGIKDPIYQKTLIETGHPKRGDIVVFKYPEDPKLDYIKRAVGLPGDKVTYDP
VSKELTIQPGCSSGQACENALPVTYSNVEPSDFVQTFSRRNGGEATSGFFEVPKNETKENGIRLSERKETLGDVTHRILT
VPIAQDQVGMYYQQPGQQLATWIVPPGQYFMMGDNRDNSADSRYWGFVPEANLVGRATAIWMSFDKQEGEWPTGLRLSRI
GGIH

>P02724 GpA Glycophorin-A (Human)
MYGKIIFVLLLSEIVSISASSTTGVAMHTSTSSSVTKSYISSQTNDTHKRDTYAATPRAHEVSEISVRTVYPPEEETGER
VQLAHHFSEPEITLIIFGVMAGVIGTILLISYGIRRLIKKSPSDVKPLPSPDTDVPLSSVEIENPETSDQ
```

Supplementary Figure 2. **cBcl2s and control protein sequences. a,** The panel includes the sequences (in a FASTA format) of Bcl2, BclXL, Bax, Bak, Bik, and Bmf. The TMD in each protein is shown in bold. **b,** The panel includes the sequences of the control proteins used in this manuscript, Tomm20 (T20), Tomm22 (T22), Lep and GpA. The TMDs are shown in bold. In those cases where the protein has two TMDs, only the one used as control is marked. The Uniprot code of each protein is shown alongside its name and the species to which it belongs.
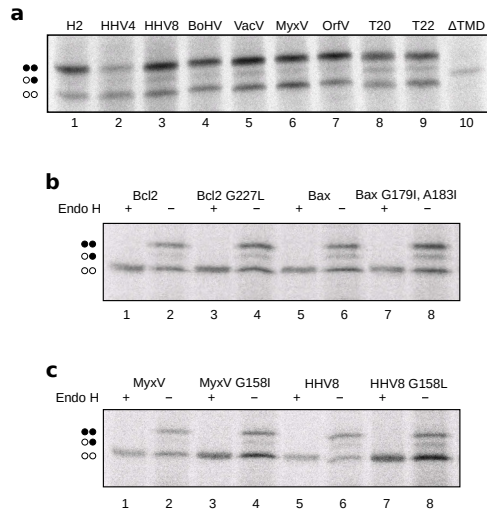
Supplementary Figure 3. **Insertion of vBcl2 Ct hydrophobic sequences in ER-derived microsomes**. **a**, In vitro translation of Lep variants where H2 has been replaced by the Ct hydrophobic region of the indicated vBcl2 proteins. As a translocation control, we used a Lep variant where H2 has been removed (ΔTMD), ΔGexp=1.50 kcal/mol. Absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots (n=3). **b** and **c**, In vitro protein translation in the presence of ER-derived microsomes. After translation Lep chimeras bearing the TMD of Bcl2, Bcl2 G227L, Bax and Bax G179I, A183I, MyxV, MyxV G158I, HHV8, or HHV G158L were treated with (+) or without (-) of Endoglycosydase H (EndoH) (n=3).
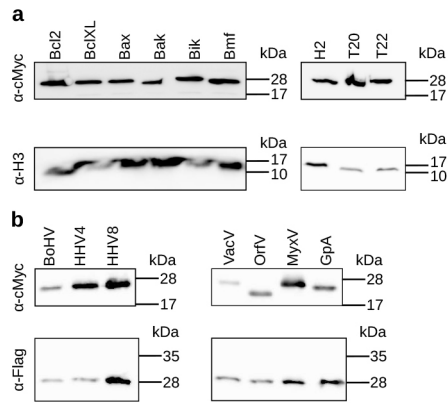
Supplementary Figure 4. **BiFC WB quantification**. **a**, WB analysis of VN chimeras bearing the TMD of Bcl2, BclXL, Bax, Bak, Bik, Bmf, H2, T20 or T22 using an anti-c-Myc antibody. Histone 3 (H3) was used as a loading control (n=3). **b**, WB quantification of VN chimeras bearing the TMD of BoHV, HHV4, HHV8, VacV, OrfV, MyxV, or GpA using an anti-c-Myc antibody (n=3). Cells were co-transfected with red fluorescent protein bearing a flag tag to facilitate protein loading quantification.
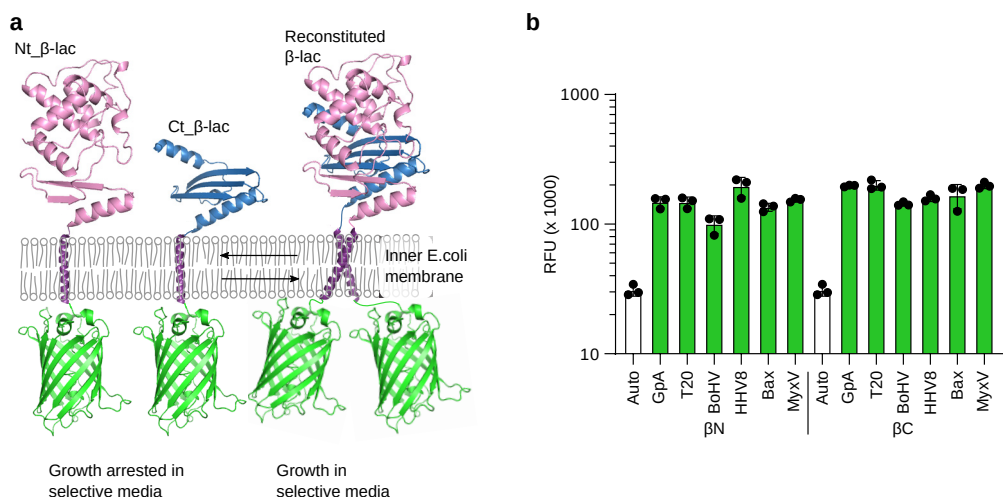
Supplementary Figure 5. **BlaTM assay in E. coli membranes**. **a**, A schematic representation of the BlaTM assay. The position of the Nt and Ct ends of the β-lactamase (in pink and blue respectively), the TMD (in purple), and the GFP (in green) are shown. βN and βC fragments (including the indicated TMDs) were expressed in E.coli in the presence of ampicillin at increasing concentrations. After 16 hours, the OD544 of the culture was measured and normalized based on the GpA homo-dimer values. The strength of the TMDs interaction is proportional to the LD50 of the antibiotic. **b**, Expression profile of all the BlaTM chimeras used in this work. To ensure that differences in the protein levels were not the source of the observed variations in LD50, we measured the GFP-derived fluorescence of all BlaTM chimeras. The mean and standard deviation fluorescence (Relative Fluorescence Units, RFU) of three independent experiments are shown (green bars) (n=3). Individual values are represented with solid dots.

Supplementary Figure 6. **Localization of TMD homo-oligomers**. **a**, **b**, and **c**, The localization of TMD-TMD homo-oligomers was analyzed by fluorescent confocal microscopy (n=4). DAPI staining is shown in blue, the BIFC signal for each of the homo-oligomers in green, and the cellular marker in red (Endoplasmatic reticulum (ER), Plasma membrane (PM), Mitochondria (Mito)). The right column of each panel shows the co-localization of BiFC and the corresponding cellular marker in yellow (visible only when the images are merged).

Supplementary Figure 7. **Hetero-oligomerization of vBcl2s TMDs and control sequences**. **a-d**, Relative fluorescence for the oligomerization of T20 TMD with vBcl2 or cBcl2 TMDs. c and d, Hetero-oligomerization of vBcl2s TMDs with T22 or Lep H2 TMDs. The bars show the mean and standard deviation of at least three independent experiments (n>3). The TMD included in the VFP chimeras (VN or VC) is indicated below each bar. Solid dots represent the results of individual experiments. The GpA TMD homo-dimer was used as a positive control and as the normalization value (dotted line). As negative controls, we used the interaction of each partner in the hetero-oligomers with T20 TMD (i.e., X/T20 and T20/Y for the X/Y interaction). An interaction was considered only if the fluorescence values were significantly higher (two-tailed homoscedastic t-test) than those of the two negative controls. No statistical differences were found.

```
Virus        Host            Host scientific
HHV4         Human           Homo sapiens
HHV8         Human           Homo sapiens
BoHV         Bovine          Bos taurus
VacV         Bovine/Horse    Bos taurus/Equus caballus
MyxV         Rabbit          Oryctolagus cuniculus
OrfV         Sheep/Goat      Ovis aries/Capra hircus
```

Bcl2
```
>Homo sapiens
WLSLKTLLSLALVGACITLGAYL
>Bos taurus
WLSLKALLSLALVGACITLGAYL
>Equus caballus
WLSLKALLSLALVGACITLGAYL
>Oryctolagus cuniculus
WVSLKTLFSLALIGACITLGAYL
>Capra hircus
WLSLKALLSLALVGACITLGAYL
```

```
Oryctolagus      WVSLKTLFSLALIGACITLGAYL
Homo             WLSLKTLLSLALVGACITLGAYL
Bos              WLSLKALLSLALVGACITLGAYL
Equus            WLSLKALLSLALVGACITLGAYL
Capra            WLSLKALLSLALVGACITLGAYL
                 *:***:*:****:**********
```

BclXL
```
>Homo sapiens
FNRWFLTGMTVAGVVLLGSLF
>Bos taurus
FNRWFLTGMTVAGVVLLGSLF
>Equus przewalskii†
FNRWFLTGMTVAGVVLLGSLF
>Oryctolagus cuniculus
FNRWFLTGMTVAGVVLLGSLF
>Ovis aries
FNRWFLTGMTVAGVVLLGSLF
```

```
Homo             FNRWFLTGMTVAGVVLLGSLF
Bos              FNRWFLTGMTVAGVVLLGSLF
Equus            FNRWFLTGMTVAGVVLLGSLF
Oryctolagus      FNRWFLTGMTVAGVVLLGSLF
Ovis             FNRWFLTGMTVAGVVLLGSLF
                 *********************
```
† The sequence from Equus przewalskii was used as a substitute of Equus caballus

Bax
```
>Homo sapiens
TWQTVTIFVAGVLTASLTIW
>Bos taurus
TWQTVTIFVAGVLTASLTIW
>Equus caballus
TWQTVTIFVAGVLTASLTIW
>Oryctolagus cuniculus
TWQTLTILGAGVLTASLTIW
>Capra hircus
TWQTVTIFVAGVLTASLTIW
>Ovis aries
TWQTVTIFVAGVLTASLTIW
```

```
Homo             TWQTVTIFVAGVLTASLTIW
Bos              TWQTVTIFVAGVLTASLTIW
```

```
Equus           TWQTVTIFVAGVLTASLTIW
Capra           TWQTVTIFVAGVLTASLTIW
Ovis            TWQTVTIFVAGVLTASLTIW
Oryctolagus     TWQTLTILGAGVLTASLTIW
                ****:**: ***********
```

Bak
```
>Homo sapiens
ILNVLVVLGVVLLGQFVVRRFF
>Bos taurus
IKSVAIVLAVVLLGQFVVRRFF
>Equus caballus
IRNVLIVLAVVLLGQYVVRRFF
>Oryctolagus cuniculus
ILTVLAALAVVAFCQFVVRR
>Capra hircus
IKNVAIVLAVVLLGQFVVRRFF
>Ovis aries
IKNVAIVLAVVLLGQFVVRRFF
```

```
Oryctolagus     ILTVLAALAVVAFCQFVVRR--
Homo            ILNVLVVLGVVLLGQFVVRRFF
Equus           IRNVLIVLAVVLLGQYVVRRFF
Bos             IKSVAIVLAVVLLGQFVVRRFF
Capra           IKNVAIVLAVVLLGQFVVRRFF
Ovis            IKNVAIVLAVVLLGQFVVRRFF
                * .*  .*.** : *:****
```

Bik
```
>Homo sapiens
LLALLLLLALLLPLLSGGLHLLL
>Equus caballus
LALSVLLLLVLLLGWGLHLL
```

```
Homo       LLALLLLLALLLPLLSGGLHLLL
Equus      -LALS-VLLLLVLLLGWGLHLL-
            ***  :* **: **. *****
```

Bmf
```
>Homo sapiens
NQNRVWWQILLFLHNLALNG
>Bos taurus
NRNRMWWQILLFLHNVALNG
>Oryctolagus cuniculus
NRNRVWWQILLFLHNLALNG
>Capra hircus
NRNRMWWQILLFLHNVALNG
>Ovis aries
NRNRMWWQILLFLHNVALNG
```

```
Bos             NRNRMWWQILLFLHNVALNG
Capra           NRNRMWWQILLFLHNVALNG
Ovis            NRNRMWWQILLFLHNVALNG
Homo            NQNRVWWQILLFLHNLALNG
Oryctolagus     NRNRVWWQILLFLHNLALNG
                *:**:**********:****
```

Supplementary Figure 8. **Alignment of cBcl2s TMDs.** The TMD sequence alignment for Bcl2, BclXL, Bax, Bak, Bik and, Bmf. The input sequences (in FASTA format) and the results of each alignment are shown. Additionally, the host affected by each of the viruses is indicated. The sequences of the human proteins were used as a reference. Sequences were obtained from the Uniprot and NCBI databases (December 2019). The alignment was done with Clustal Omega (EMBL-EBI).
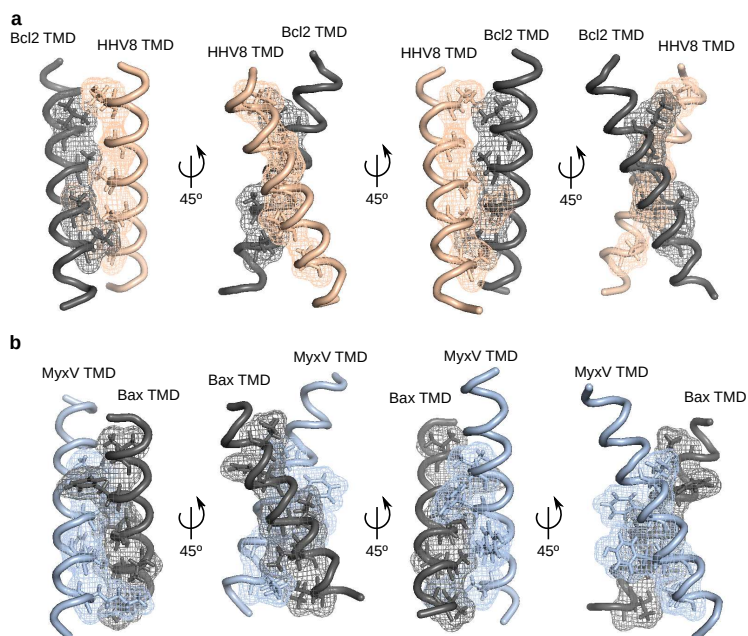
Supplementary Figure 9. **Models for the Bcl2-HHV8 and Bax-MyxV intramembrane interactions**. **a**, Model of a putative dimer between Bcl2 and HHV8 TMDs, obtained with PredDIMER. **b**, Model of a putative dimer between Bcl2 and MyxV TMDs, obtained with PredDIMER.

VN

| VC	Bcl2	BclXL	Bax	Bak	Bik	Bmf	HHV4	HHV8	BoHV	VacV	MyxV	OrfV	T20	T22	H2
Bcl2	8E-06	n.d.	n.d.	n.d.	n.d.	n.d.	1E-05	0.0297	0.0039	0.0055	0.0005	0.0185	na	n.d.	n.d.
BclXL	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.929	0.229	0.810	0.0166	0.394	0.0445	na	n.d.	n.d.
Bax	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.879	0.672	0.557	0.0029	0.339	0.0124	na	n.d.	n.d.
Bak	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.388	0.486	0.315	0.0168	0.0113	0.0305	na	n.d.	n.d.
Bik	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	**0.0164**	0.514	0.654	0.0015	0.854	0.0397	na	n.d.	n.d.
Bmf	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.734	0.401	0.074	0.217	0.739	0.877	na	n.d.	n.d.
HHV4	0.0368	0.0403	0.418	0.626	0.0205	0.841	0.001	n.d.	n.d.	n.d.	n.d.	n.d.	na	0.136	0.237
HHV8	0.0009	0.0184	0.162	0.691	0.0003	0.0316	n.d.	0.0386	n.d.	n.d.	n.d.	n.d.	na	0.189	0.921
BoHV	0.179	0.753	0.626	0.837	0.416	0.151	n.d.	n.d.	0.1423	n.d.	n.d.	n.d.	na	0.444	0.817
VacV	0.005	0.0244	0.814	0.0164	0.0019	0.884	n.d.	n.d.	n.d.	0.0082	n.d.	n.d.	na	0.109	0.960
MyxV	0.0105	0.664	0.0367	0.0475	0.334	0.664	n.d.	n.d.	n.d.	n.d.	0.0028	n.d.	na	0.024	0.686
OrfV	0.0142	0.0222	0.501	0.597	0.497	0.536	n.d.	n.d.	n.d.	n.d.	n.d.	0.0002	na	0.053	0.778
T20	na	na	na	na	na	na	na	na	na	na	na	na	0.810	na	na
T22	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.138	0.682	0.123	0.781	0.261	0.465	na	0.394	n.d.
H2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.801	0.251	0.658	0.932	0.748	0.714	na	n.d.	6E-07

Supplementary Figure 10. **Analysis of viral-host TMD-TMD interactions by BiFC**. The result of all the TMD-TMD interactions assayed in HEK 293T cells using the BiFC assay summarized in a matrix format. Fluorescence values (RFU) for each of the indicated BiFC combinations significantly higher than those of their respective negative controls are highlighted in green (two-tailed homoscedastic t-test, p-value < 0.05). Only the highest of the two corresponding p-values is shown. Absence of interaction is shown in white. Grey boxes indicate combinations not done (nd) or not applicable (na).
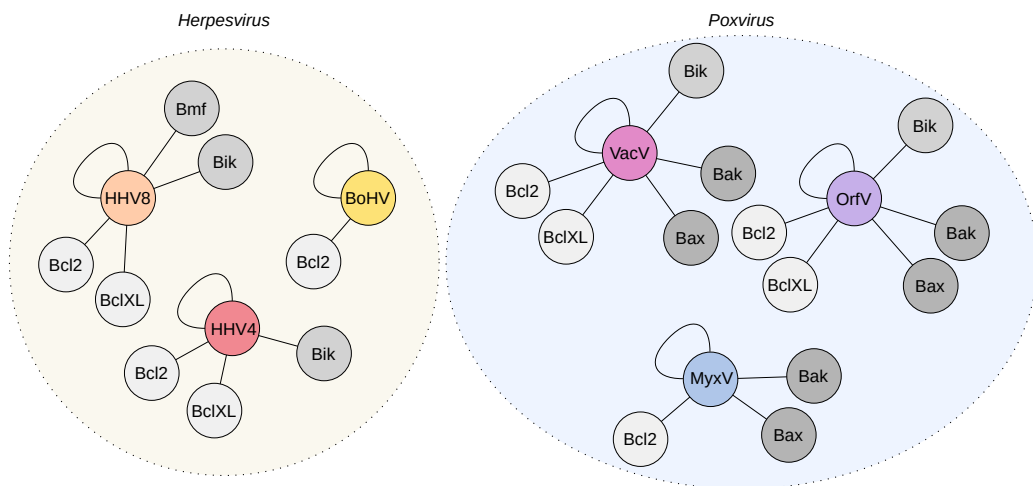
Supplementary Figure 11. **TMD-TMD interactions network of the vBcl2**. A network representation of the host-cell interaction of vBcl2 TMDs. The figure includes the results of the BiFC and BlaTM assays. Solid lines represent interactions, while TMDs are represented by nodes. The colored areas highlight the viral families, with poxviruses in light blue and herpesviruses in light yellow.
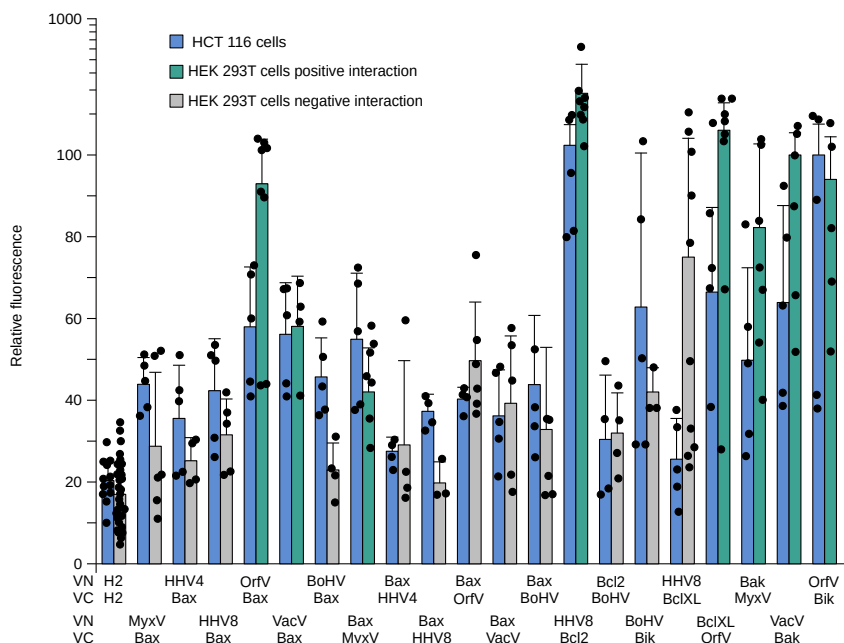
Supplementary Figure 12. **BiFC assay in HCT 116 cells**. Bar graph showing the relative fluorescence of the tested homo-oligomers in the BiFC assay in HCT 116 cells (blue bars). The mean and standard deviation of at least five independent experiments are shown (n>5). The corresponding value in HEK 293T cells is shown to facilitate comparison. In HEK 293T cells, those interactions that were statistically above their controls are shown in green, while grey bars denote the absence of intramembrane interaction (see Fig. 3).
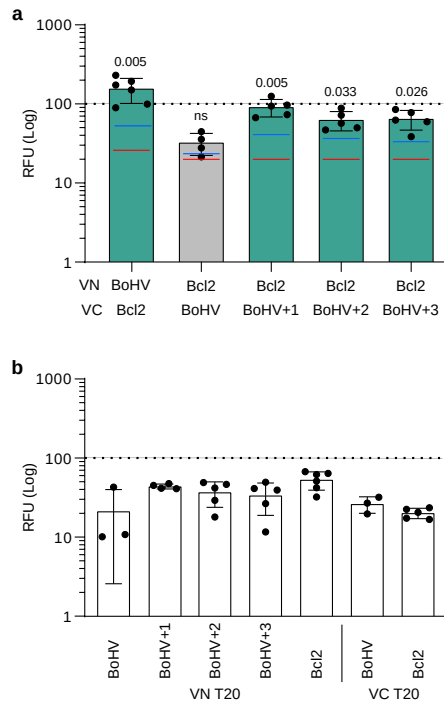
Supplementary Figure 13. **Analysis of the interaction between Bcl2 and BoHV TMDs**. **a** and **b**, Relative fluorescence (RFU) for the hetero-oligomerization between VN-Bcl2 and VC-BoHV, VC-BoHV+1, VC-BoHV+2, or VC-BoHV+3 (a). The results for the VN-BoHV/VC-Bcl2 oligomerization are included for comparison (previously shown in Fig 3.). The mean and standard deviation of at least three independent experiments are shown (n>3). Solid dots represent the results of individual experiments. The TMD included in the VFP chimeras (VN or VC) is indicated below each bar. The GpA TMD homo-dimer was used as a positive control and as the normalization value (dotted line). The interactions of each partner in the hetero-oligomers with T20 TMD were used as a negative control (b). To facilitate comparison, blue (VN/T20 TMD) and red (VC/T20 TMD) lines within each bar in panel a indicate the fluorescence of the corresponding controls. An interaction (highlighted in green) would be considered only if the corresponding fluorescence is significantly higher than those of the negative controls (two-tailed homoscedastic t-test). The highest of the two p-values is shown, ns non significant.
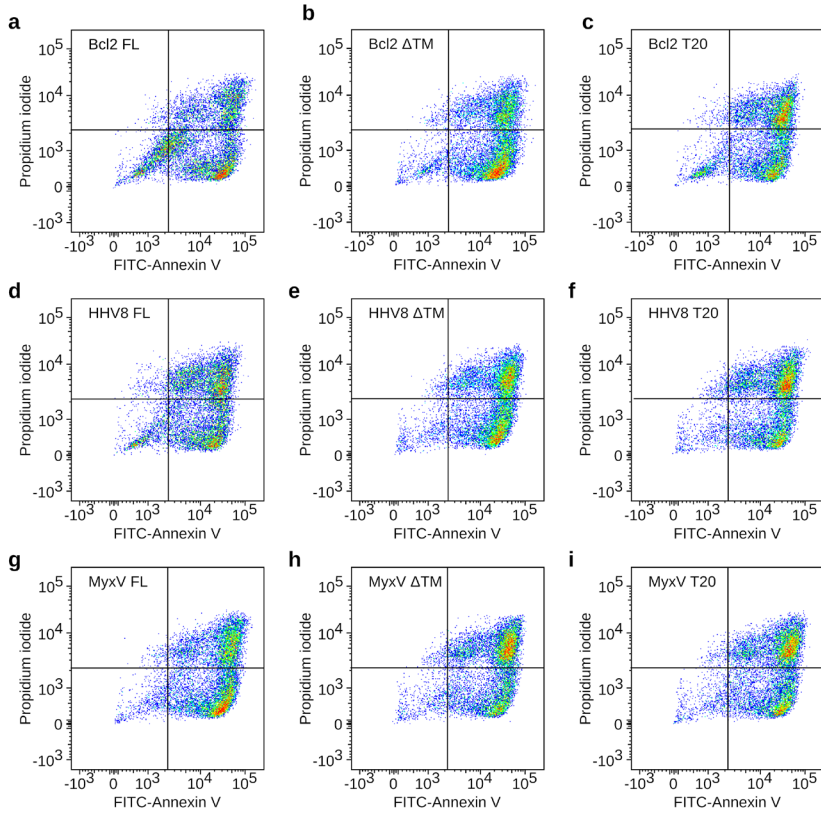
Supplementary Figure 14. **Flow cytometry analysis of vBcl2 TMD in apoptotic control**. **a-i**. Cells were transfected with Bcl2, HHV8 and MyxV with (FL) or without the TMD (ΔTMD), or with the TMD substituted by the TMD of T20 (T20), and treated with doxorubicin as an apoptotic stimulus. The percentage of cells in healthy, early apoptotic, late apoptotic, or necrotic states was measured by flow cytometry using propidium iody staining and phosphatidil serine labeling (FITC-Annexin V). Panels a-i show a representative plot for each of the indicated treatments.
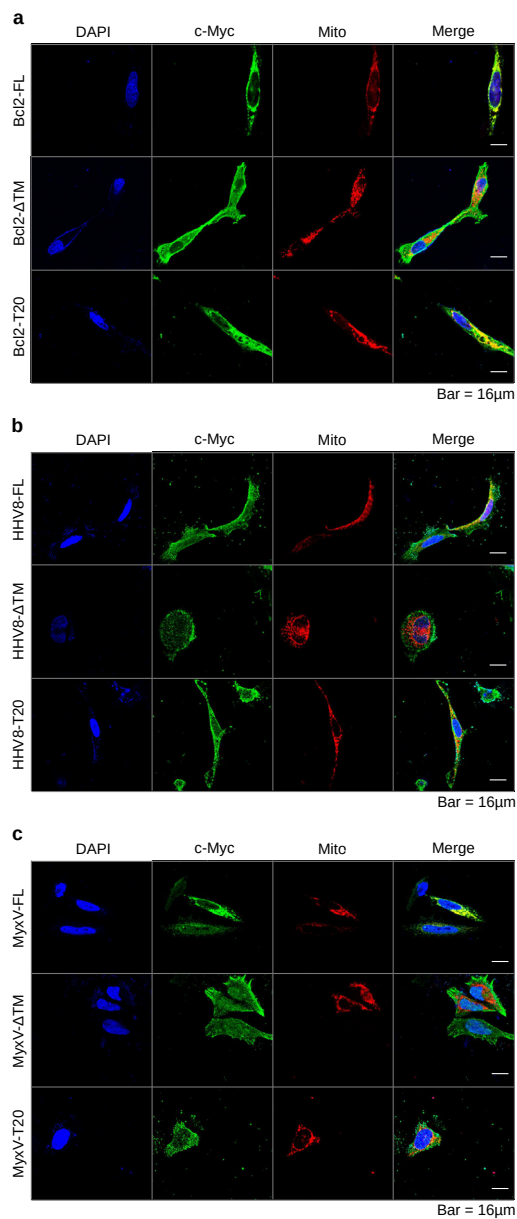
Supplementary Figure 15. **Sub-cellular localization of Bcl2, HHV8 and, MyxV**. **a-c**, Sub-cellular localization of Bcl2, HHV8 and MyxV with (FL) or without (ΔTMD) the TMD, or with the TMD replaced by the TMD of T20. vBcl2 proteins were transfected in HeLa cells and immunostained using an anti-c-Myc antibody (green) (n=4). To monitor mitochondrial localization, a marker (mCherry-T20) was co-transfected (red). DAPI staining is shown in blue. The right column of each panel shows the co-localization of the vBcl2 protein and the mitochondrial marker in yellow (visible only when the images are merged).
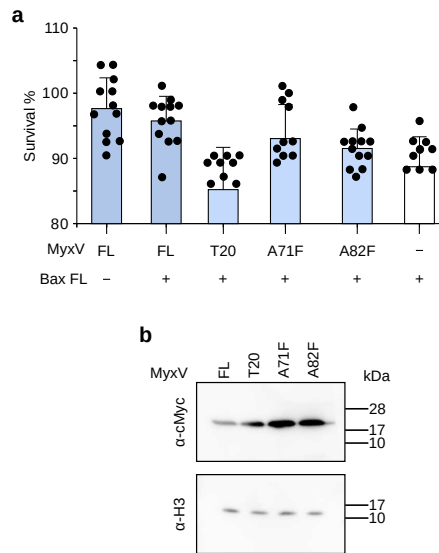
Supplementary Figure 16. **Influence of MyxV binding groove and TMD interactions on apoptotic control**. **a**, Cells were transfected with an empty plasmid (-), MyxV FL and T20, or MyxV bearing substitutions in alanines 71 or 82 to phenylalanines (A71F and A82F, respectively) and co-transfected with (+) or without (-) Bax FL as an apoptosis stimulus. Survival percentage mean and standard deviation of six replicates from two independent experiments are shown. **b**. Western blot analysis of protein levels. Histone 3 (H3) was used as a loading control (n=3).

Supplementary Table 1. **Used primers**

Name	sequence
vBcl2 gene string1_F	AATGTCACCCTAAGCTTCTTCGTGA
vBcl2 gene string1_R	AGAAGAAACCAACCTCGAGTCGGTA
vBcl2 gene string2_F	GTACAATGTCACCCTAGATCTTTTG
vBcl2 gene string2_R	TGTGAAGAAGAAACCAACGGTACCT
vBcl2 gene string3_F	TACAATGTCACCCTAAGCTTACCTG
vBcl2 gene string3_R	TGAAGAAGAAACCAACGGTGCCTCT
HHV4_Lep_F	GCATACTAGTTTTAGCTGGACTTTG
HHV4_Lep_R	GCATGGTACCCCATATAACTAACAG
HHV8_Lep_F	GCATACTAGTAGAATGACAGCGCTA
HHV8_Lep_R	GCATGGTACCCATCGCGACCGCTGC
BoHV4_Lep_F	GCATACTAGTCGCTCCAGATGGTTA
BoHV4_Lep_R	GCATGGTACCTCTCGCCACACCCAC
VacV_Lep_F	GCATACTAGTCGCGAGTACCTGAAA
VacV_Lep_R	GCATGGTACCCAGGGTTTTGTAGGT
MyxV_Lep_F	GCATACTAGTATCAGCGTGTACCTG
MyxV_Lep_R	GCATGGTACCTCTGTACCACTTCAG
OrfV_Lep_F	GCATACTAGTCGCGCCCTGGGCGTG
OrfV_Lep_R	GCATGGTACCGCGCAGCACGCGCAC
Bcl2_Lep_F	GCATACTAGTTGGCTGTCTCTGAAG
Bcl2_Lep_R	GCATGGTACCCAGATAGGCACCCAG
Bax_Lep_R	GCATACTAGTACCTGGCAGACCGTG
Bax_Lep_F	GCATGGTACCCCAAATGGTCAGGCT
T20_Lep_R	GCATACTAGTGCCATCGCCGCCGGC
T20_Lep_F	GCATGGTACCGAAGTAGATGCAGTA
T22_Lep_R	GCATACTAGTGCCCTGTGGATCGGC
T22_Lep_F	GCATGGTACCGGTCTCGAAGACGAC
HHV8_G158L_Lep_F	AGTAGAATGACAGCGCTATTGTTAAGCATTGCATTATTGGCCAC
HHV8_G158L_Lep_R	GTGGCCAATAATGCAATGCTTAACAATAGCGCTGTCATTCTACT
T7 ampli TNT Lep F	ATAGTATAATACGACTCACTATAGGAAACCACCATGGCGAATTCCACC
ampli Lep TNT end R	CTATTAATGGATGCCGCC
Bak_BIFC_F1 annealing	GGCCGCATCCTGAACGTGCTGGTGGTTCTGGGTGTGGTTCTGTTGGGCCAG
Bak_BIFC_F2 annealing	TTTGTGGTACGAAGATTCTTCAAATCATGAGC
Bak_BIFC_R1 annealing	AACCACACCCAGAACCACCAGCACGTTCAGGATGC
Bak_BIFC_R2 annealing	GGCCGCTCATGATTTGAAGAATCTTCGTACCACAAACTGGCCCAACAG
Bax_BIFC_F1 annealing	GGCCGCACGTGGCAGACCGTGACCATCTTTGTGGCGGGAGTGCTCACC
Bax_BIFC_F2 annealing	GCCTCGCTCACCATCTGGAAGAAGATGGGCTGAGC
Bax_BIFC_R1 annealing	CGCCACAAAGATGGTCACGGTCTGCCACGTGC
Bax_BIFC_R2 annealing	GGCCGCTCAGCCCATCTTCTTCCAGATGGTGAGCGAGGCGGTGAGCACTCC
T20_VN_in-fusion_F	GCGGGAGTAGCGGCCGCGCCATCGCCGCCGGCGTG
T20_VN/VC_in-fusion_R	TGGATCCCCGCGGCCGCTCAGAAGTAGATGCAGTA
T20_VC_in-fusion_F	AGCAGAAGAGCGGCCGCGCCATCGCCGCCGGCGTG
T22_VN_in-fusion_F	GCGGGAGTAGCGGCCGCGCCCTGTGGATCGGC
T22_VN/VC_in-fusion_R	TGGATCCCCGCGGCCGCTCAGGTCTCGAAGACGAC
T22_VC_in-fusion_F	AGCAGAAGAGCGGCCGCGCCCTGTGGATCGGC
HHV8_FL_BIFC_F	AAGGAAAAAAGCGGCCGCGACGAGGACGTTTTGCC
HHV8_DTM_BIFC_R	TTTTCCTTTTGCGGCCGCTCATCCCCTTCTTCTGGTAAG
HHV8_FL_BIFC_ R	TTTTCCTTTTGCGGCCGCTCATCTCCTGCTCATCGCG
HHV8_T20_BIFC_PCR1 R1	CACACGCCGGCGGCGATGGCTTCTCCGTCCCCTTC
HHV8_T20_BIFC_PCR2 R2	GCAGTAGCCGATGAACAGGGCGCCGCACACGCCGG
HHV8_T20_BIFC_PCR3 R3	TTGCGGCCGCTCAGAAGTAGATGCAGTAGCCGATG
HHV8_T20_BIFC_PCR4 R4	TTTTCCTTTTGCGGCCGCTCA
Bcl2_FL_BIFC_F	AAGGAAAAAAGCGGCCGC ATGGCGCACGCTGGGAG
Bcl2_FL_BIFC_R	TTTTCCTTTTGCGGCCGCTCACTTGTGGCTCAGAT
Bcl2_DTM_BIFC_R	TTTTCCTTTTGCGGCCGCTCAAGGAGAAATCAAACAGAG
Bcl2 G227L_R	CAGTTTGGCCCTGGTGCTAGCTTGCATCAC
Bcl2 G227L_F	CAGGGTGATGCAAGCTAGCACCAGGGCCAA
MyxV_G158I_BIFC_R	GCGATACCATTTCAGAATGATATACGCCACAAAGCCCACC
MyxV_G158I_BIFC_F	GGTGGGCTTTGTGGCGTATATCATTCTGAAATGGTATCGC
MyxV_G158I_pVOTE_F	GGGCTTCGTGGCCTACATCATCCTGAAGTGGTAC
MyxV_G158I_pVOTE_R	GTACCACTTCAGGATGATGTAGGCCACGAAGCCC
MyxV A82F_R	GCAGGGTCAGGGTGAACAGCTTCACGCTGG
MyxV A82F_F	CCAGCGTGAAGCTGTTCACCCTGACCCTGC
MyxV A71F_R	GCTGGCGTCTCTCAGGAAGAAGGTCAGGGTGTTC
MyxV A71F_F	GAACACCCTGACCTTCTTCCTGAGAGACGCCAGC

Name	sequence
Bax_G179I_BlaTM_R	GCTCGCGGTCAGCACGATCGCCACAAAAATGGTC
Bax_G179I_BlaTM_F	GACCATTTTTGTGGCGATCGTGCTGACCGCGAGC
Bax_A183I_BlaTM_R	TCCAAATGGTCAGGCTTATGGTCAGCACGCCCGCC
Bax_A183I_BlaTM_F	GGCGGGCGTGCTGACCATAAGCCTGACCATTTGGA
Bax_G179I/A183I_BlaTM_F	GGCGATCGTGCTGACCATAAGCCTGACCATTTGGA
Bax_G179I/A183I_BlaTM_R	TCCAAATGGTCAGGCTTATGGTCAGCACGATCGCC
Bax_FL_G179I_F	CGAGGCGGTGAGCACTATCGCCACAAAGATGGTC
Bax_FL_G179I_R	GACCATCTTTGTGGCGATAGTGCTCACCGCCTCG
Bax_FL_A183I_F	CCAGATGGTGAGCGAGATGGTGAGCACTATCGCC
Bax_FL_A183I_R	GGCGATAGTGCTCACCATCTCGCTCACCATCTGG
VC-BoHV+3_BIFC_F	GGCCGCCCAAACTTGAAGCTTCCTAGCCCCCGCTCCAGATG
VC-BoHV+3_BIFC_R	CATCTGGAGCGGGGGCTAGGAAGCTTCAAGTTTGGGCGGCC
VC-BoHV+2_BIFC_F	GGCCGCCCAAACTTGAAGCTTAGCCCCCGCTCCAGATG
VC-BoHV+2_BIFC_R	CATCTGGAGCGGGGGCTAAGCTTCAAGTTTGGGCGGCC
VC-BoHV+1_BIFC_F	GGCCGCCCAAACTTGAAGCTTCCCCGCTCCAGATG
VC-BoHV+1_BIFC_R	CATCTGGAGCGGGGAAGCTTCAAGTTTGGGCGGCC
KpnI-cMyc-HHV8_pCAGGS_F	GCATGGTACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGGACGAGGACGTTTTGCCTGG
XhoI-HHV8 FL pCAGGS_R	TGCTAGCTCGAGTTATCTCCTGCTCATCGC
XhoI-HHV8 DTM pCAGGS_F	TGCTAGCTCGAGTTATCCCCTTCTTCTGGTAAG
HHV8_T20_pCAGGS_PCR1 R1	ACGCCGGCGGCGATGGCTCCCCTTCTTCTGGTAAG
HHV8_T20_pCAGGS_PCR2 R2	GTAGCCGATGAACAGGGCGCCGCACACGCCGGCGG
HHV8_T20_pCAGGS_PCR3 R3	ACGTCTCGAGTCAGAAGTAGATGCAGTAGCCGATG
In-fusion _pCAGGS _T20_R	GATCTGCTAGCTCGATCAGAAGTAGATGCAGTAGCCG
KpnI-cMyc-Bcl2_pCAGGS_F	GCATGGTACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGGCGCACGCTGGGAGAAC
XhoI-Bcl2 FL_pCAGGS_R	TGCTAGCTCGAGTTACTTGTGGCTCAG
XhoI-Bcl2 DTM pCAGGS_R	TGCTAGCTCGAGTCAGGAGAAATCAAACAG
Bcl2_T20_pCAGGS_R1	ACGCCGGCGGCGATGGCGGAGAAATCAAACAGAGG
cMyc-MyxV_pCAGGS_F	ACCATGGAGCAGAAGCTGATCAGCGAGGAGGACCTGAGCAGACTGAAGACC
MyxV-T20 _pCAGGS_PCR1 R	ACGCCGGCGGCGATGGCCTTGCAGCCGCCGCTTCT
MyxV-FL_pCAGGS R	GATCTGCTAGCTCGATCAGGTGCCTCTGTACCACT
MyxV-DTM-pCAGGS_R	TCACTTGCAGCCGCCGC
T20_pCAGGS_IF_R(new)	CTCGAGCATGCCCGGTCAGAAATAAATGCAATA
VAcV_FL_infusion_R	GATCTGCTAGCTCGAGATCATGTACTTCAGGGTCTTG
InFusion VacV FL reverse	GATCTGCTAGCTCGACTAGCTCGAGATCATGTACTTCAGG
InFusion VacV DTM reverse	GATCTGCTAGCTCGACTAGGTGGAGGTGTTGGGG
pVote cmyc to pCAGGSKpnI_F	GCTCATCGATGCATGCCACCATGGAGCAGAAGCTGATCAGC
pVote to pCAGGS KpnI_R	CTCGAGCATGCCCGGATTCCCGGGAGCTCGAGC
MyxV_G158I_IF_pCAGGS_R1	CACTTCAGGATGATGTAGGCCACGAAGCCCACCAC
MyxV_G158I_IF_pCAGGS_R2	CTCGAGCATGCCCGGTCAGGTGCCTCTGTACCACTTCAGGATGAT
cMyc_NdeI_pVOTE2_F	AACACGATAATCATAATGGAGCAGAAGCTGATCAGC
Bcl2_FL_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCACTTGTGGCTCAGATAGGC
Bcl2_DTM_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCAGGAGAAATCAAACAGAGGCCG
T20_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCAGAAGTAGATGCAGTAGCCG
T22_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCAGGTCTCGAAGACGACGG
MyxV_FL_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCAGGTGCCTCTGTACCACTTCAGG
MyxV_DTM_NdeI_pVOTE2_R	GGGAGCTCGAGCATATCACTTGCAGCCGCCGCTTCTCACG
HHV8_FL_NdeI_pVOTE2_R	GGGAGCTCGAGCATATTATCTCCTGCTCATCGCGACC
HHV8_DTM_NdeI_pVOTE2_R	GGGAGCTCGAGCATATTCCCCTTCTTCTGGTAAGCACC
VacV_FL_NdeI_pVOTE2_R	GGGAGCTCGAGCATAGATCATGTACTTCAGGGTCTTGT
VacV_DTM_NdeI_pVOTE2_R	GGGAGCTCGAGCATAGGTGGAGGTGTTGGGGATTG
HHV4_anneal_BlaTM_F1	CTAGCTTTAGCTGGACCCTGTTTCTGGCGGGCCTG
HHV4_anneal_BlaTM_F2	ACCCTGAGCCTGCTGGTGATTTGCAGCTATCTGTTTATTGG
HHV4_anneal_BlaTM_R1	GATCCCAATAAACAGATAGCTGCAAATCACCAGCA
HHV4_anneal_BlaTM_R2	GGCTCAGGGTCAGGCCCGCCAGAAACAGGGTCCAGCTAAAG
BoHV_anneal_BlaTM_F1	CTAGCTGGCTGTTTCCGATGTTTGCGATTAGCGGC
BoHV_anneal_BlaTM_F2	CTGGTGCTGACCGTGGGCGTGGCGCGCGG
BoHV_anneal_BlaTM_R1	GATCCCGCGCGCCACGCCCACGGTCAGCACCAGGC
BoHV_anneal_BlaTM_R2	CGCTAATCGCAAACATCGGAAACAGCCAG
VacV_anneal_BlaTM_F1	CTAGCGAATATCTGAAACTGATTGGCATTACCGCG
VacV_anneal_BlaTM_F2	ATTATGTTTGCGACCTATAAAACCCTGGG
VacV_anneal_BlaTM_R1	GATCCCCAGGGTTTTATAGGTCGCAAACATAATCG
VacV_anneal_BlaTM_R2	CGGTAATGCCAATCAGTTTCAGATATTCG
OrfV_anneal_BlaTM_F1	CTAGCGCGCTGGGCGTGGGCGCGGTGGTGGCGGGC
OrfV_anneal_BlaTM_F2	GTGGGCATGCTGCTGCTGGGCGTGCGCGTGCTGCGCGG

Name	sequence
OrfV_anneal_BlaTM_R1	GATCCCGCGCAGCACGCGCACGCCCAGCAGCAGCA
OrfV_anneal_BlaTM_R2	TGCCCACGCCCGCCACCACCGCGCCCACGCCCAGCGCG
T20_anneal_BlaTM_F1	CTAGCGCGATTGCGGCGGGCGTGTGCGGCGCGCTG
T20_anneal_BlaTM_F2	TTTATTGGCTATTGCATTTATTTTGG
T20_anneal_BlaTM_R1	GATCCCAAAATAAATGCAATAGCCAATAAACAGCGC
T20_anneal_BlaTM_R2	GCCGCACACGCCCGCCGCAATCGCG
HHV4_BlaTM_InFusion_F	TGCTAATCGAGCTAGCTTTAGCTGGA
HHV4_BlaTM_InFusion_R	GCCAGTTTGTGGATCCCGATAAATAAA
BoHV_BlaTM_InFusion_F	TGCTAATCGAGCTAGCTGGTTATTTCCCA
BoHV_BlaTM_InFusion_R	GCCAGTTTGTGGATCCCTCTCGCCACACCCA
VacV_BlaTM_InFusion_F	TGCTAATCGAGCTAGCGAGTACCTGAAG
VacV_BlaTM_InFusion_R	GCCAGTTTGTGGATCCCCAGGGTCTTGT
T20_BlaTM_InFusion_F	TGCTAATCGAGCTAGCGCCATCGCCGCCG
T20_BlaTM_InFusion_R	GCCAGTTTGTGGATCCCGAAGTAGATGC
OrfV_BlaTM_InFusion_F	TGCTAATCGAGCTAGCGCCCTGGGCGTG
OrfV_BlaTM_InFusion_R	GCCAGTTTGTGGATCCCGCGCAGCACGC
Seq BIFC CMV-promoter	CGCAAATGGGCGGTAGGCGTG
Seq BIFC Ct Reverse	CAGCCAGCCGCGGCC
Seq pCAGGS F	CCTTCTTCTTTTTCCTACAGC
Seq pCAGGS R	GATGTCCCCATAATTTTTGGCAGAGGG
Seq pVECTOR2 F	GGCCCCCCGAACCACGGGGAC
Seq BlaTM_R	CTGAACACCATAGGTCAGGGTGGTAACCAGG
Seq_N_BlaTM_F	TTATCTACACGACGGGGAGTCAGGCAACTATG
Seq_C_BlaTM_F	GAAGCCATACCAAACGACGAGCGTGACACC

4.4. Chapter 4: Membrane protein interaction design

Computational design of BclxL inhibitors that target transmembrane domain interactions

Gerard Duart[1], Assaf Elazar[2], Jonathan J. Weinstein[2], Laura Gadea-Salom[1], Juan Ortiz-Mateu[1], Sarel J. Fleishman[2], Ismael Mingarro[1], Luis Martinez-Gil[1‡]

[1]Department of Biochemistry and Molecular Biology, Institut de Biotecnologia i Biomedicina BIOTECMED, Universitat de València, 46100, Burjassot, Spain.
[2]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel.
[‡]Corresponding author: luis.martinez-gil@uv.es

Abstract

Most pathological events associated with apoptosis resistance are related to the overexpression of anti-apoptotic Bcl2 family members. These events include newly diagnosed cancers as well as acquired therapy resistance. Among the anti-apoptotic proteins, BclxL has a predominant role in several forms of cancer, and different strategies for inhibiting it have been considered, including antisense oligonucleotides, proteolysis targeting chimeras, and BH3 mimetics. However, attempts at BclxL-targeting therapies have foundered because of a lack of efficacy or adverse effects. Here, we explored the intramembrane protein–protein interactions of BclxL and their role in its anti-apoptotic function. Based on our findings, we computationally designed sequences capable of specifically sequestering the transmembrane domain of BclxL. In cancer-derived cell lines we demonstrate that the inhibitor cancels the anti-apoptotic effect of BclxL without toxic side effects. Our results represent the first steps in the development of a unique type of BclxL inhibitor that could supplement or substitute for some current cancer treatments. Moreover, our approach may trigger the development of a new generation of inhibitors targeting interactions between transmembrane domains.

Introduction

Apoptosis, the main mechanism of controlled cell death, is an evolutionarily conserved cellular process that occurs in response to various physiological and pathological situations[1]. Morphological and biochemical hallmarks of apoptosis include cell shrinkage, nuclear DNA fragmentation, and membrane blebbing. Most of these changes result from the activation of a set of cysteine proteases known as caspases[2]. Given the importance of caspase activation for cell fate, understanding the activation of these proteins is crucial for controlling apoptosis if its deregulation occurs.

Activation of apoptotic caspases represents the last step of two signaling routes, the so-called extrinsic and intrinsic pathways. The extrinsic pathway is initiated by ligand binding to a death receptor of the tumor necrosis factor receptor family. Stimulation of the receptor causes caspase 8 activation, which in turn activates caspase 3[3]. The intrinsic pathway, which originates in the mitochondria, begins with the release of apoptogenic factors such as cytochrome c or endonuclease G from these organelles. This release also prompts caspase 3 activation and ultimately apoptosis[3]. Permeabilization of the mitochondrial outer membrane (MOM) leads to the release of apoptogenic factors and is primarily regulated by the B-cell lymphoma 2 (Bcl2) protein family[4]. Based on the function of either preventing or promoting MOM permeabilization, Bcl2 members are classified as anti-apoptotic and pro-apoptotic proteins. Anti-apoptotic members include Bcl2, BclxL, Mcl1, and BclW, and pro-apoptotic members of the family consist of effector proteins such as Bak, Bax, and Bok, and the BH3-only proteins, which are subdivided into sensitizers (Bad, Noxa, Bmf, Hrk, Bik) and activators (Bim, Bid, and Puma)[5].

Bcl2 members have a variable number of Bcl2 homology (BH) domains. In addition, many Bcl2 family proteins have a transmembrane domain (TMD) in the carboxyl-terminal (Ct) end that facilitates the insertion of the protein into the target lipid bilayer[6]. Members of the Bcl2 protein family can interact with each other, forming homo- and hetero-oligomers[7–10]. These protein–protein interactions (PPIs) are part of an important regulatory mechanism of MOM permeabilization and thus programmed cell death. In healthy cells,

anti-apoptotic Bcl2 members inhibit activation of pro-apoptotic proteins through direct interaction or by sequestering BH3-only proteins[4]. Upon an apoptotic stimulus, BH3-only and pro-apoptotic proteins are released and free to induce MOM permeabilization. Interactions among Bcl2 family members have been thought to occur through soluble domains, especially BH domains[11]. However, recent findings suggest that their TMDs also participate in these PPIs[12-17] and that these intramembrane interactions are crucial for apoptotic control[14]. Nevertheless, intramembrane PPIs are poorly understood in comparison with their soluble counterparts, despite increasing evidence of their importance, not only because of apoptosis control but also because of the intrinsic difficulty of working with membrane proteins. As such, they have rarely been explored as an alternative for protein control[18,19].

Virtually all pathological events associated with apoptosis deregulation can be linked to apoptosis resistance. Over-expression of anti-apoptotic Bcl2 members is common in newly diagnosed cancers and associated with resistance to several treatments, including various chemotherapeutic agents and γ-irradiation[20-23]. Conversely, loss or down-regulation of pro-apoptotic members has been reported in many tumor types[5,24]. These observations have sparked the development of drugs targeted at anti-apoptotic Bcl2 members to control aberrant survival.

Among the anti-apoptotic proteins, the Bcl-2–like protein 1, better known as BclxL, displays relevant functions in several forms of cancer. In melanoma, BclxL participates in many of the hallmarks of this aggressive form of cancer, such as preventing cells from executing apoptosis and inducing drug resistance, cell migration and invasion, and angiogenesis[25]. Because of the relevance of BclxL in the progression of cancer, different strategies have been considered to inhibit it[25]. These strategies include antisense oligonucleotides, Proteolysis Targeting Chimeric (PROTAC) molecules, and BH3 mimetics. The BH3 mimetics were designed to disrupt the PPIs that keep pro-apoptotic members at bay[26]. Many excellent reviews have extensively covered the use of these compounds[25-28], but briefly, BH3 mimetics have shown promising results at different stages of research and trials and at a clinical level. However, some caveats remain unresolved.

Commonly used BH3 mimetics rely mostly on Bcl2 inhibition because of the adverse effect of targeting BclxL[29], so non-Bcl2–dependent forms of cancer cannot be efficiently treated with the current mimetics. Even in cases in which prevailing BH3 mimetics are adequate, BclxL-associated resistance can emerge. For this reason, new and less adverse BclxL inhibitors are needed. These novel therapeutics could be used as an alternative to current BH3 mimetics or in combination with current treatments and may represent a significant step in the development of targeted, personalized therapy.

Here, we explore the intramembrane PPIs of BclxL and their role in the anti-apoptotic function of this Bcl-2 family member. Using this information, we computationally designed an inhibitor capable of selectively sequestering the TMD of BclxL. Our designed inhibitor can limit the anti-apoptotic effect of BclxL in HeLa cells using doxorubicin as an apoptotic stimulus. Our results constitute the first steps in the development of a new type of transmembrane BclxL inhibitor that may supplement or substitute some current cancer treatments. Moreover, we hope that this work sparks the development of a new generation of inhibitors targeting interactions between transmembrane domains.

Results

The TMD of BclxL establishes interactions with pro- and anti-apoptotic Bcl2 members

Bcl2 and Mcl-1 can establish intramembrane interactions with other members of the Bcl2 family[12,13]. Furthermore, to block apoptosis, viral analogs of Bcl2 proteins must interact through the TMD with cellular Bcl2 members[14]. We sought to explore whether the BclxL TMD can be used to establish interactions with other pro- and anti-apoptotic Bcl2 members. To assess these potential intramembrane contacts, we employed BLaTM, a genetic tool designed to qualitatively and quantitatively study TMD–TMD interactions in *E. coli*[30]. Briefly, the tested TMDs are fused to either the Nt or the Ct end of a split β-lactamase (βN and βC, respectively) and the enhanced green fluorescent protein (eGFP) (Figure 1a). Additionally, the chimeras include the pelB cleavable signal peptide, which directs the protein to the inner bacterial membrane and determines its topology, ensuring periplasmic

localization of the β-lactamase. In the inner membrane, an efficient TMD–TMD interaction facilitates the reconstitution of the β-lactamase and thus the growth of bacteria in selective media (ampicillin) (Figure 1a). In this assay, the LD_{50} of the antibiotic serves as an indicator of the strength of the assayed TMD–TMD interaction, while the eGFP-derived fluorescence allows for rapid quantification of protein levels. As a positive control and for normalization purposes, we used the TMD of glycophorin A (GpA), a hydrophobic segment that can form noncovalent homodimers within the membrane[31–33]. The non-oligomerizing TMD of the mitochondrial protein Tomm20 (T20) was used as a negative control for membrane overcrowding and stochastic interactions[14]. Of note, bacteria can grow only in the presence of ampicillin when the β-lactamase is reconstituted in the periplasm, i.e., only if the tested regions are properly inserted in the bacterial inner membrane. Therefore, the BLaTM assay also indicates the insertion potential of the tested regions. Using this approach, we tested the homo-oligomerization of BclxL TMD and its hetero-oligomerization with the TMD of anti-apoptotic Bcl2, and the TMDs of pro-apoptotic (Bax and Bak) and BH3-only (Bik and Bmf) Bcl2 members. Our results indicated that BclxL TMD forms a weak homo-oligomer (Figures 1b and 1c). Furthermore, we identified transmembrane hetero-oligomers with Bcl2, Bax, and Bak.

The reporter signal resulting from a TMD–TMD interaction depends not only on the sequences of the interacting TMDs and thus their inherent affinity, but also on the orientation of the interacting surfaces of the TMDs in relation to their accompanying signaling domains[14,30]. To confirm that the lack of interaction between the TMD of BclxL and the TMDs of Bik and Bmf was not the result of misaligned reporter domains[30] we tested opposing combinations of BLaTM chimeras, i.e., βN-Bik/βC-BclxL and βN-Bmf/βC-BclxL (Figure 1d). Despite the change in the experimental set-up, we observed no interaction between the BclxL TMD and the TMDs of Bik or Bmf. Of note, expression levels, measured using GFP fluorescence, were comparable for all BLaTM chimeras (Supp. Figure 1).
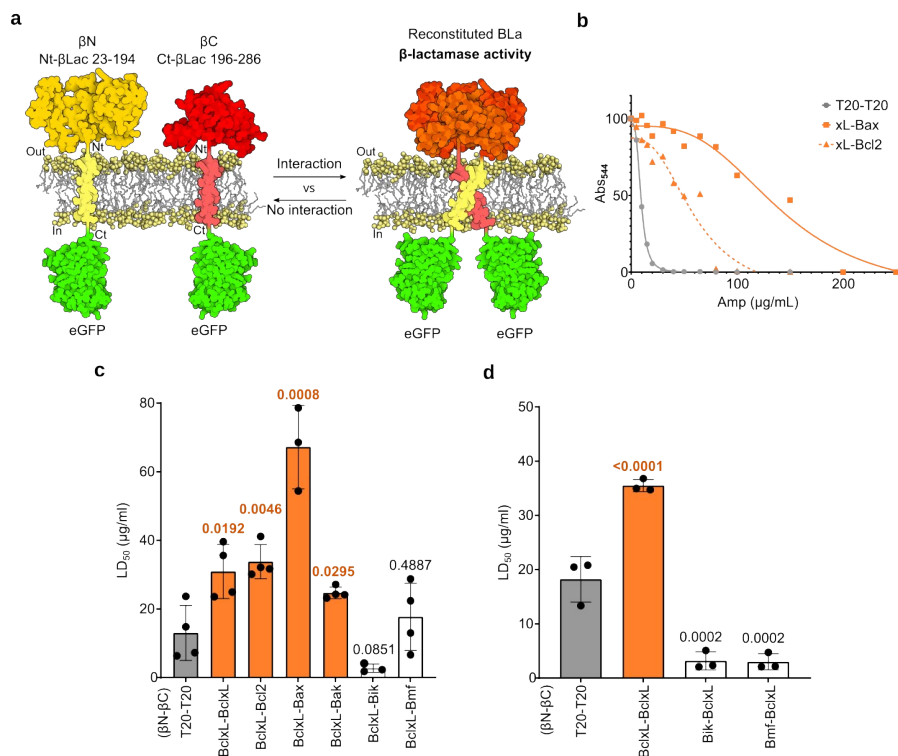
Figure 1. BlcxL's TMD interactions in biological membranes. **a**. Schematic representation of the BLaTM assay. The β-Lactamase was split in two non-active fragments (Nt-βLac and Ct-βLac). Each of them was fused to a TMD and the eGFP in C-terminus generating the βN and βC chimeras respectively. A TMD–TMD interaction approximates the Nt-βLac and Ct-βLac fragments and facilitate the reconstitution of the β-Lactamase structure and activity (ampicillin resistance). The residues constituting the Nt-βLac and Ct-βLac halves are specified above the protein representation. **b**. In the BLaTM assay, the strength of a TMD–TMD interaction correlates with resistance to ampicillin. The panel shows representative examples of the dose-response curves from which the ampicillin LD_{50} was calculated. The dose-response curves corresponding with the interactions between the TMDs of BclxL and Bax (orange, squares), BclxL and Bcl2 (orange, triangles), and the TMD of T20 are shown. **c**. and **d**. The βN and βC chimeras bearing the TMD of the indicated proteins were co-expressed in *E. coli* and the resulting ampicillin LD_{50} was measured. The βN T20-βC T20 homodimer was used as a negative control (gray), and the βN GpA-βC GpA homodimer was used as a positive control and normalization value across experimental replicates (LD_{50}~100 μg/mL). The normalized means and standard deviations of at least three independent experiments (n ≥ 3) are shown. The individual value for each experiment is represented by a solid dot. An interaction (highlighted in orange) was considered if the observed LD_{50} was

significantly higher (two-tailed homoscedastic t-test, p-value < 0.05) than the negative control (gray bar). P-values are indicated above the corresponding bar.

Next, to corroborate the interaction capabilities of the BclxL TMD in eukaryotic cells, we used a bimolecular fluorescent complementation (BiFC) aproach[34], adapted for the study of intramembrane interactions[12,14,35]. Briefly, the tested TMDs were fused to a split Venus fluorescent protein (VFP), either to its N-terminus (VN) or its C-terminus (VC), neither of which is fluorescent. Interaction of the TMDs brought together the VN and VC ends, reconstituting the VFP structure and restoring its fluorescence (Figure 2a). As noted, the TMD of GpA was used as a positive control and normalization value across experimental replicates, and the TMD of T20 was used as a negative control. Our results indicated that the BclxL TMD can homo-oligomerize in eukaryotic membranes. Furthermore, using BiFC, we detected hetero-oligomers with the TMDs of Bcl2, Bak, and Bik (Figure 2b). We also tested the VN Bax-VC BclxL and VN Bmf-VC BclxL combinations of BiFC chimeras (Figure 2c). Despite the change in the experimental set-up, no interaction was observed between the TMD of BclxL and the TMDs of Bax or Bmf. Of note, western blot analysis indicated that the BiFC chimera bearing the Bax TMD was consistently expressed at lower levels, which could explain the differences seen with the BlaTM assay (Figure 2d). For a more comprehensive visualization of the aforementioned interactions, we also have included a network representation (Supp. Figure 2).

*BclxL intramembrane interactions are crucial for its anti-apoptotic role*
Next, to investigate whether these newly found TMD–TMD interactions are necessary for BclxL control of cellular apoptosis, we transfected HeLa cells with BclxL with or without the TMD (BclxL-FL and BclxL-ΔTMD, respectively). Additionally, we included a chimera in which the TMD of BclxL was replaced by the TMD of T20 (BclxL-T20). All of these constructs included a Nt c-myc tag to facilitate detection. As a negative control, cells were transfected with an empty plasmid (Empty). Once transfected, cells were treated with doxorubicin to induce apoptosis[14,36] and the percentage of surviving cells 16 h post-treatment was calculated based on Trypan blue staining (Figure 3a and b). Our results indicated that when doxorubicin is used as a cell death stimulus, BclxL requires the TMD to block apoptosis.

Furthermore, the substitution of this TMD by the TMD of T20 impaired the anti-apoptotic function of BclxL, suggesting that intramembrane interactions of BclxL with other Bcl2 members are necessary for BclxL's anti-apoptotic function. Western blot analysis confirmed comparable expression levels for the BclxL-FL, BclxL-ΔTMD, and BclxL-T20 variants (Figure 3c).
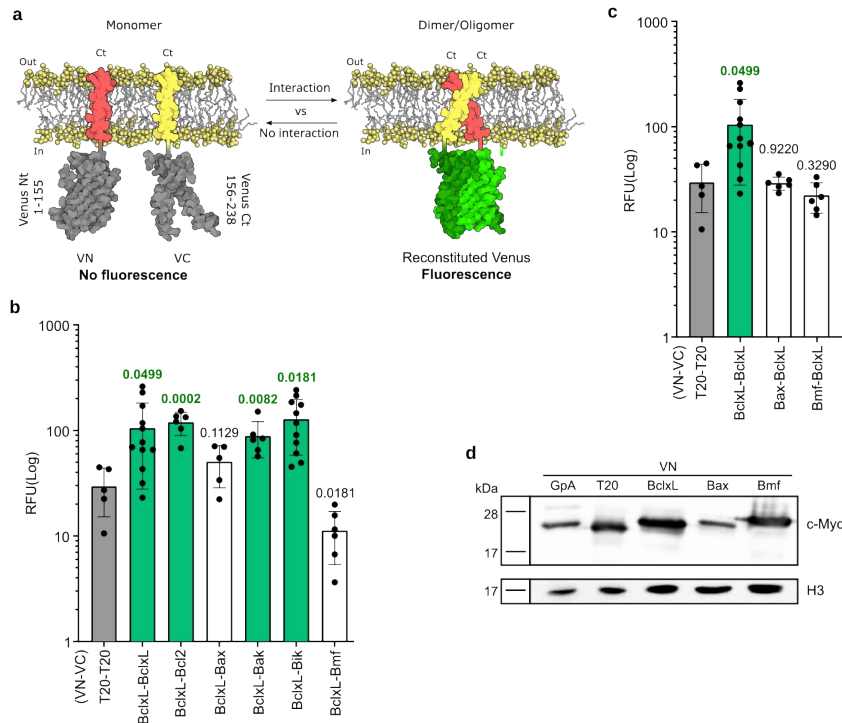
Figure 2. Interactions of the BclxL TMD in eukaryotic membranes. **a**. Schematic representation of the Bimolecular Fluorescent Complementation (BiFC) assay. The two non-fluorescent fragments of a split Venus fluorescent protein (Venus Nt and Venus Ct; both in gray) are fused to two potential interacting TMDs (yellow and red) rendering two chimeric constructs designated respectively as VN and VC. The association of the TMDs facilitates the reconstitution of the fluorescent protein structure (green). The residues included in each VFP fragment are indicated next to the protein representation. **b** and **c** Relative fluorescence units (RFU) for the homo-oligomerization of the TMD of BclxL and its hetero-oligomerization with the TMD of Bcl2, Bax, Bak, Bik, and Bmf; the mean and standard deviation (n $\geq$ 5) are shown. Solid dots represent the results of individual experiments. The TMD included in each chimera (either VN or VC) is indicated below each bar. The VN T20-VC T20 combination was used as negative control (gray bar). The VN GpA-VC GpA was used as a positive control and normalization value across experiments. An

interaction (highlighted in green) was considered to have occured only if the obtained RFU was significantly higher (two-tailed homoscedastic t-test, p-value < 0.05) than the negative control, p-values are indicated above the corresponding bars. **d**. Western blot analysis of VN-GpA, VN-T20, VN-BclxL, VN-Bax, and VN-Bmf protein levels. Histone 3 (H3) was used as a loading control. Experiments were done in triplicates.
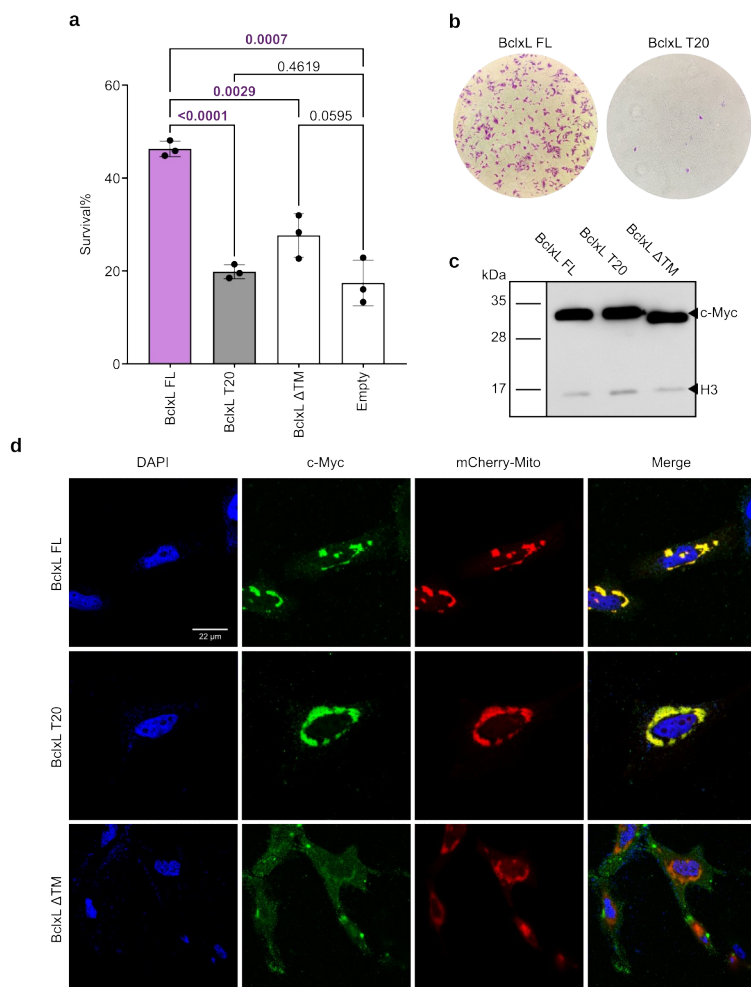
Figure 3. The importance of the BclxL TMD in apoptotic control. **a**. Survival of doxorubicin-treated cells. HeLa cells were transfected with BclxL with or without the TMD (BclxL-FL and BclxL-ΔTMD, respectively), we also included a chimera in which the TMD of BclxL was replaced by the TMD of T20 (BclxL-T20). Next, cells were treated with doxorubicin [15 µM] final, and the percentage of surviving cells

16 h post-treatment was calculated based on Trypan blue staining. The survival percentage means and standard deviations of three independent experiments are shown (n = 3). Solid dots represent the results of individual experiments. Transfection with an empty plasmid (Empty) was used as a negative control. The level of significance (ordinary one-way ANOVA test with Dunnett correction, p-value < 0.05) is shown, and significant differences are highlighted in purple. **b.** Alternatively cells were transfected with Bclxl FL or BclxL T20 and treated with doxorubicin. Approximately 16 h post-treatment cells were washed, fixed and stained. Representative images are included. Images were taken using a 10x microscope objective. **c**. Western blot analysis of BclxL-FL, BclxL-ΔTM, and BclxL-T20 protein levels. Histone 3 (H3) was used as a loading control. A representative assay (n = 3) is shown. **d**. Subcellular localization of BclxL chimeras. BclxL-FL, BclxL-ΔTM, and BclxL-T20 with a c-Myc tag were transfected into HeLa cells together with a mitochondrial marker (mCherry-mito, red). After 24 h, cells were immunostained using an anti-c-Myc antibody (green). DAPI staining is shown in blue. The right column of each panel shows the co-localization of the BclxL chimeras and the mitochondrial marker (yellow), visible only when the images are merged. Experimental n = 3.

BclxL localizes primarily in the mitochondria[37–39]. Given the importance of TMDs for membrane protein sorting[40], particularly in the case of tail-anchoring proteins[41], we thought it was important to consider how deletions or substitutions in the Ct hydrophobic region of BclxL would affect its cellular localization. To analyze the subcellular location of the previously mentioned chimeras, BclxL (BclxL-FL) and the BclxL-T20 and BclxL-ΔTMD variants were expressed in HeLa cells alongside a fluorescent mitochondrial marker (mCherry-Mito, Figure 3d). The fluorescence micrographs revealed a similar subcellular distribution between BclxL-FL and BclxL-T20. Furthermore, both moieties were located at the mitochondria, as indicated by co-localization with the mitochondrial marker. BclxL-ΔTMD, however, showed a different cellular distribution. Removing the TMD left the protein in the cytosol, impeding its co-localization with the mitochondrial marker.

Computational design of BclxL inhibitors
Once we had established the importance of BclxL TMD–TMD interactions for the anti-apoptotic function, we sought to design an inhibitor that could block these intramembrane PPIs. Such an inhibitor should, in principle, block the pro-survival effect of BclxL and thus sensitize cells to an apoptotic stimulus.

Inhibitor design started with the modeling of the BclxL homo-interaction using TMHOP (Trans-membrane Homo Oligomer Predictor)[42]. TMHOP uses Rosetta symmetric all-atom ab initio fold-and-dock simulations in an implicit membrane environment to predict thousands of low-energy conformations based on the energy function described by Weinstein et al.[42]. This energy function relies on the empirical measurement of amino acid insertion propensities into the bacterial inner membrane[43]. It encodes a lipophilicity term for each amino acid, depending on each amino acid's exposure to the lipid and its distance from the cytosolic or the extracellular milieu. TMHOP generates several models for the desired interaction (Figure 4a). Based on structural characteristics and associated Rosetta energy, we selected a TMHOP model that forms a tightly packed parallel dimer. The structure of the dimer and the monomer that constitutes it can be found in Figure 4b. Roughly half of each monomer's surface area ($2389.5Å^2$) is in contact with its companion ($1244.1Å^2$). The amino acids involved in the contact between monomers are shown in Figure 4 c. The crossing angle of the dimer was exactly 48°, with the shortest distance (4.1Å) between the N atoms of Met at position 9. To avoid confusion, we numbered the amino acids based on their position within the TMD.

The selected model was fed into the FuncLib design algorithm[44] to generate higher affinity binders that could serve as inhibitors. The FuncLib algorithm uses Rosetta design calculations to enumerate combinations of tolerated amino acid substitutions at specific positions. It then relaxes each combination using whole-protein minimization (based on the Rossetta membrane energy function[42]) and ranks these combinations by energy. In addition to the atomistic design steps, FuncLib uses information from multiple-sequence alignments to eliminate mutations that are not commonly observed in homologs. This method has been successfully applied to dramatically improve enzyme catalytic rates[44], binding affinity[45], and antibody stability and affinity[46], but it has not been applied previously to membrane proteins. Additionally, single-span TMDs tend to self-associate[35], whereas our design objective was an inhibitor that would target BclxL specifically and exhibit minimal self-association propensities. To do so we

computationally modeled each design to obey the following rules. 1. Positive selection of a heterodimer (Non-symmetric FuncLib; ΔΔG<+1 Rosetta energy units; R.e.u.); 2. Negative selection for a homodimer (symmetric FuncLib; ΔΔG>+5 R.e.u.).
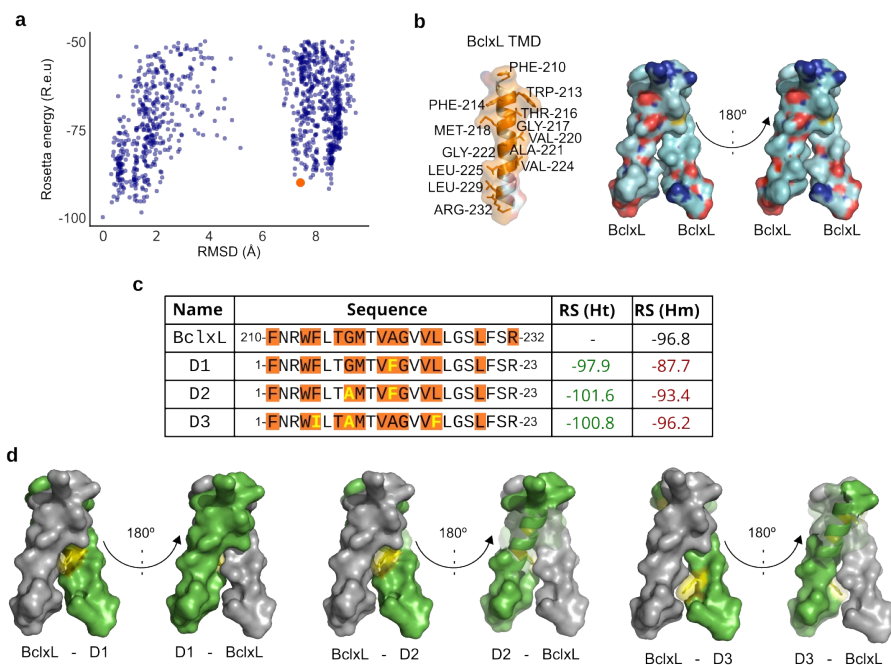
Figure 4. Design of BclxL TMD inhibitors. **a**. Scatter plot distribution for the BclxL TMD homodimer modeling. Dots represent each of the models for the BclxL TMD's homodimers. RMSD is calculated from the lowest energy model. The selected model is highlighted in orange. b. Structural representations of the selected model. The structure of the monomer and the dimer (frontal and 180º turn views) are shown. The monomer representation features the amino acids on the interacting surface, numbering based on the BclxL sequence. In the dimer representation, the nitrogen and oxygen atoms are highlighted in blue and red respectively c. Sequences of the TMDs of the BclxL and the D1, D2, and D3 designs. Differences among the sequences are featured in yellow. The amino acids in the interaction surfaces of the potential dimers between the TMD of BclxL (gray) and the D1, D2, and D3 designs (green) are shown in orange. The Rosetta scores for the hetero- (RS Ht) and homodimerizations (RS Hm) are indicated. Values lower and higher than RS(Hm) for BclxL TMD are shown in green and red respectively. The positions of the amino acids in the TMD of BclxL (210-232) and in the designs (1-23) are included. d. Frontal view (and 180º turn) of the potential dimer between the TMD of BclxL (gray) and the D1, D2, and D3 designs (green). Changes in the designs over the TMD sequence of BclxL are highlighted in yellow.

The resulting potential TMD inhibitors were manually curated after visualization of the 3D structural models, and we selected three designs (designated as D1, D2, and D3). The sequences of each selected model, as well as the BclxL TMD, together with their corresponding Rosetta energy for their homo- and hetero-dimerization, are shown in (Figure 4c). Structural representations of the interactions among the designs (D1, D2, and D3) and the TMD of BclxL are given in Figure 4d. Additionally, the amino acids found in the contact surface, the contact area between monomers, and their crossing angles are shown in Supp. Figure 3.

Membrane insertion potential of D1, D2, and D3

Before testing the inhibition properties of the designed sequences, we verified that these segments could be inserted into the membrane together with BclxL. To that end, we tested the membrane insertion of the D1, D2, and D3 sequences into ER-derived microsomes using an *in vitro* transcription/translation assay. This assay, designed for an accurate and quantitative description of the membrane insertion capability of short sequences, is based on the *E. coli* leader peptidase (Lep). Lep is a Nt/Ct luminal membrane protein consisting of two TMDs (H1 and H2) connected by a cytoplasmic loop (P1) and a large C-terminal domain (P2) (Supp. Figure 4a). We used a Lep-derived system (Lep′) containing an extended Nt[47–49]. Additionally, the Lep' includes two N-glycosylation acceptor sites consisting of an Asn-X-Ser/Thr sequence, where X can be any amino acid except Pro[50]. The first acceptor site (G1) is in the extended Nt, whereas the second site (G2) is in the P2 domain. N-linked glycosylation has been extensively used as a topological reporter for more than two decades[51]. This post-translational modification occurs only in the ER lumen, the location of the active site of oligosaccharide transferase, a translocon-associated enzyme responsible for oligosaccharide transfer[52]; thus, no N-linked glycosylation occurs in polypeptide sequences spanning the membrane or facing the cytosol. In Lep', the hydrophobic region to be tested replaced the H2 domain (Supp. Figure 4a). When the tested hydrophobic region is recognized by the translocon as a TMD and inserted into the membrane, both G1 and G2 are oriented towards the ER lumen, yielding a doubly-glycosylated version of

Lep'. If the tested region is not inserted into the membrane, however, G2 stays on the cytoplasmic side and is not modified by the oligosaccharide transferase. Glycosylation of an acceptor site increases the molecular mass of the protein by ≈ 2.5 kDa relative to the observed molecular mass in the absence of membranes, allowing monitoring of the glycosylation state (one vs. two glycosylations) and thus insertion into the membrane of the tested sequence (Supp. Figure 4a right).

According to our results, all three designed sequences (D1, D2, and D3) are efficiently inserted into the membrane, as indicated by the double glycosylation pattern (Supp. Figure 4b). In this assay, for control sequences, we used the hydrophobic regions of the Bcl2 proteins Mcl1 (insertion control) and Noxa (no-insertion control)[13,53]. The experimentally obtained ΔGs (ΔG_{app}^{exp}) correlated with the *in silico* insertion potential of the designed sequences (Supp. Figure 4b), based on values calculated using the *ΔG prediction server*[54,55] with the default parameters. Of note, the ΔG_{app}^{pred} for any of the three designs was lower than the ΔG_{app}^{pred} associated with the TMD of BclxL (-0.355 kcal/mol).

Analysis of the interaction between D1, D2, or D3 with the TMD of BclxL
Next, using BLaTM, we analyzed the interactions between the TMD of BclxL and the computationally designed inhibitors D1, D2, and D3 (Figures 5a, 5b). The results of these experiments revealed that D1 can efficiently bind to the BclxL TMD but does not form homo-oligomers, as we intended in our design. Of note, the interaction between the TMD of BclxL and D1 was stronger than the homo-oligomerization of the TMD of BclxL. Although D2 and D3 did not form homo-oligomers, they did not interact with the TMD of BclxL (Figure 5b), regardless of the assayed combination (Figure 5c). Thus, we found no significant differences among the average LD$_{50}$ values associated with the BclxL TMD-D2, BclxL TMD-D3, and T20-T20 BLaTM assays. Expression levels (measured using eGFP fluorescence) were comparable for all βN and βC chimeras (Supp. Figure 5).
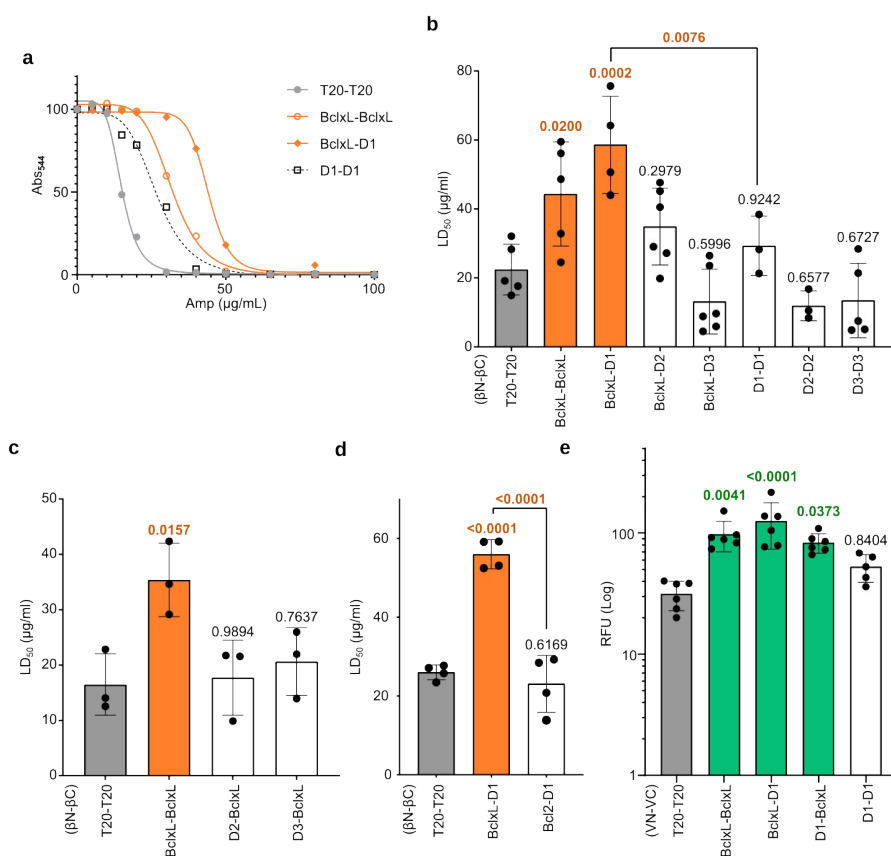
Figure 5. Interaction between the TMD of BclxL and the designed inhibitors. **a**. Representative example of ampicillin dose-response curves in the BLaTM assay. The TMD of T20 was used as negative control (gray). The TMD of GpA was used as positive control and normalization value across experiments. The BclxL-BclxL (orange circles), BclxL-D1 (orange diamonds), and D1-D1 (black dashed line, squares) interaction profiles are shown. **b**, **c**, and **d**. Interaction between the TMD of BclxL and the designed inhibitors. BlaTM chimeras (βN-βC) bearing the indicated TMD were co-expressed in *E. coli* and the resulting ampicillin LD_{50} was measured. The means and standard deviations of at least three independent experiments are shown. Individual values for each experiment are represented by solid dots. The βN T20-βC T20 homodimer was used as a negative control (gray). The βN GpA-βC-GpA homodimer was used as a positive control (not included). An interaction (highlighted in orange) was considered if the obtained LD_{50} was significantly higher (ordinary one-way ANOVA test with Dunnett correction, p-value < 0.05) than the negative control. P-values are indicated above the corresponding bar. The level of significance (ordinary one-way ANOVA test with Dunnett correction, p-value <0.05) when comparing xL-D1 vs D1-D1 is included. **e**. BiFC analysis of the interaction between the TMD of BclxL and D1 in eukaryotic membranes. The TMD included in

each chimera (VN or VC) is indicated. The RFU mean and standard deviation of at least five independent experiments are shown. Solid dots represent the results of individual experiments. The TMD of T20 was used as negative control (gray). The TMD of GpA was used as a positive control and normalization value across experiments. An interaction (highlighted in green) was considered if the obtained RFU was significantly higher (two-tailed homoscedastic t-test, p-value < 0.05) than the negative control (gray bar). P-values are indicated above the corresponding bar (significantly higher in bold green).

Next, we investigated the specificity of the observed interaction by challenging D1 with the TMD of Bcl2, another anti-apoptotic protein. To do so, we once again used the BLaTM approach and tested the BclxL TMD-D1 and Bcl2 TMD-D1 interactions side by side. We detected no interaction between D1 and the Bcl2 TMD (Figure 5d). Thus, any effect of D1 on cell survival would most likely arise from its interaction with the BclxL TMD. The similar expression levels of all BLaTM chimeras used in this assay (Supp. Figure 5) indicated that the observed differences were not the result of different protein concentrations within the cells.

Additionally, we used the BiFC assay to ensure that the interaction between D1 and the BclxL TMD was maintained in eukaryotic membranes (Figure 5e). The results indicated that D1 could efficiently bind to the TMD of BclxL but did not form homo-oligomers. Of note, the D1 and BclxL TMD hetero-interaction was identified regardless of the combination of BiFC chimeras analyzed.

D1 and D2 share a single amino acid change over the TMD sequence of BclxL, specifically Ala12Phe (Figure 4c). Of note, this is the only difference between D1 and the TMD of BclxL. In light of these data, we decided to test whether other amino acid substitutions in the same position would facilitate the interaction with the BclxL TMD while also avoiding homo-interactions. First, we analyzed which substitutions at position 12 of the TMD of BclxL would allow insertion of the resulting sequence into the mitochondrial membrane. For this purpose, we used the ΔG prediction server[54,55] (Supp. Figure 6a). The amino acid substitutions that kept the ΔGpred of the resulting segment below zero (indicative of membrane insertion) were included in the BLaTM assay for assessment of their homo-oligomerization

and hetero-oligomerization (with the TMDs of BclxL and Bcl2) properties. Only one of the tested substitutions (Ala12Cys) showed an interaction profile similar to D1. The Ala12Cys substitution facilitated interaction with the BclxL TMD while avoiding the formation of homo-oligomers or off-target PPIs with the TMD of Bcl2, emphasizing the specificity of our computational design (Supp. Figure 6b and c).

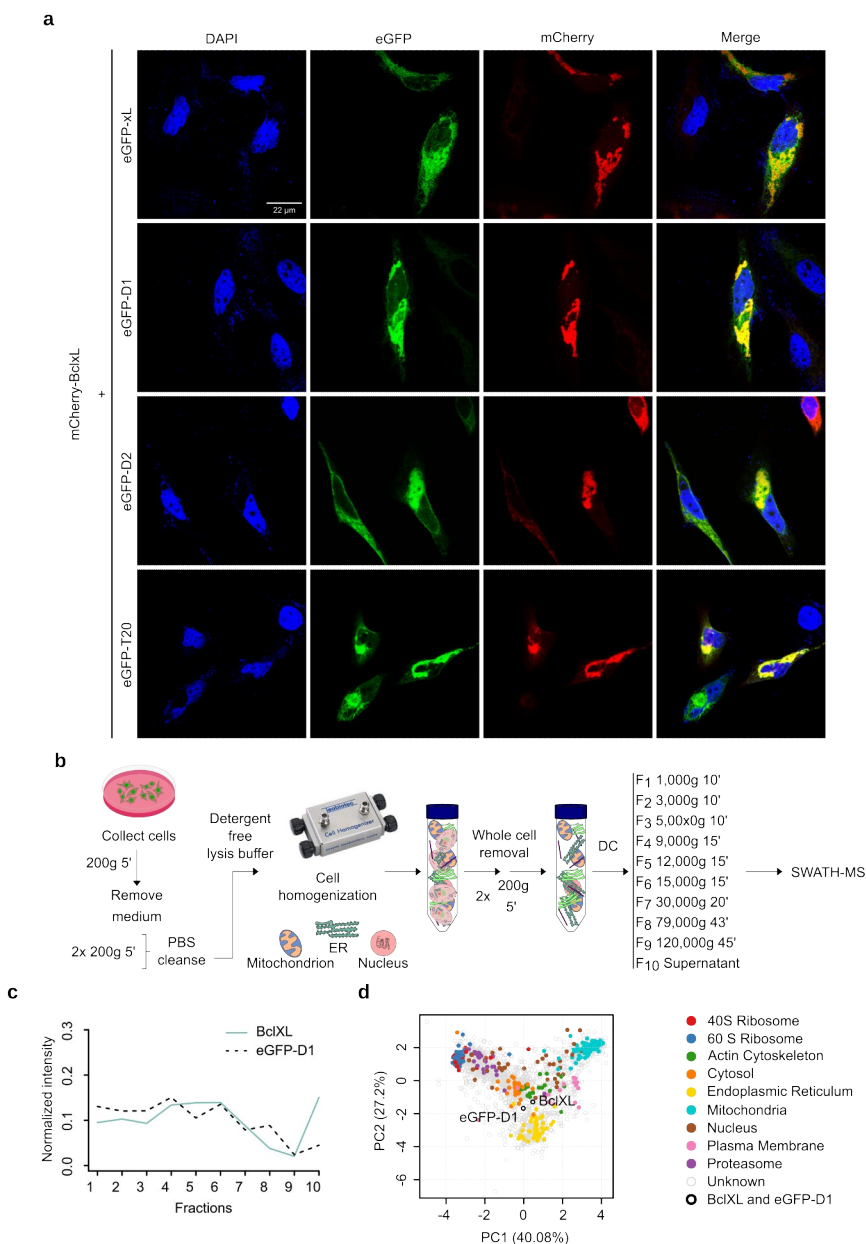*Subcellular localization of BclxL inhibitors*

To inhibit the anti-apoptotic effect of BclxL, the designed sequences must be located in the same cellular compartment where BclxL is found. To test their location, we fused D1 and D2 sequences to the Ct of the eGFP (eGFP-D1 and eGFP-D2) and expressed this construct in HeLa cells together with BclxL attached to the fluorescent protein mCherry (mCherry-BclxL). D3 was excluded from the assay because of the poor interaction with BclxL TMD observed in the BLaTM assay. We intentionally fused the designed sequences at the Ct of the eGFP to mimic the topology of a protein of the Bcl2 family and consequently its localization[56]. Next, we analyzed the subcellular distribution of these chimeras by confocal fluorescence microscopy (Figure 6a). Additionally, chimeras bearing the TMD of BclxL or T20 were included in the assay as controls (eGFP-xL and eGFP-T20, respectively). As expected, eGFP-xL and mCherry-BclxL had the same subcellular distribution. mCherry-BclxL and eGFP-T20 were also co-localized in HeLa cells. Analysis of the locations of eGFP-D1 and eGFP-D2 showed a strong co-localization with mCherry-BclxL (Figure 6), indicating a similar cellular distribution for the D1 and D2 inhibitors and BclxL. We did not observe co-localization of the mCherry-BclxL with eGFP bearing no TMD (Supp. Figure 7).

To ensure that D1 and BclxL coexist in the same cellular compartment and thus that D1 could block the anti-apoptotic role of BclxL, we performed a localization assay based on organelle differential utracentrifugation[57]. Briefly, cells expressing either BclxL or eGFP-D1 were homogenized in a detergent-free medium to avoid organelle content release. Next, the cell lysate was divided into 10 fractions by sequential differential centrifugation, as previously described[57] (Figure 6b). We then analyzed the protein content

in each fraction by sequential window acquisition of all theoretical mass spectra (SWATH-MS)[58,59]. Data analysis and visualization, as well as protein localization, were done with a dedicated open-source R package[60,61]. BclxL and eGFP-D1 had a similar distribution profile, suggesting a similar subcellular localization (Figure 6c and 6d, Supp. Figure 8). Minor differences were found at fraction 10, suggesting that Bclx exists partially as a cytosolic protein[62–64]. A comparison of D1 and BclxL profiles with a set of organelles markers revealed no distinct protein localization for either of these proteins (Supp. Figure 8 and 9). Of note, similar results have been obtained previously for Bcl2[65], Bax[57], and mouse BclxL[66]. A list of the organelle markers together with the SAWTH-MS data can be found in Data Set 1.

D1 and D2 inhibit the anti-apoptotic effect of BclxL
Finally, we tested the anti-apoptotic effect of D1 and D2. HeLa cells were transfected with BclxL alongside the eGFP-T20, eGFP-D1, eGFP-D2, or eGFP-xL chimeras. As a control, we used cells that did not receive BclxL or any of the chimeras and transfected them with an empty plasmid (Empty) to keep the amount of transfected DNA constant across all samples. After transfection, cells were treated with doxorubicin to induce apoptosis. The cells that received eGFP-T20 or eGFP-xL plus BclxL could block doxorubicin-induced apoptosis (Figures 7a and b). Remarkably, transfection of eGFP-D1 eliminated the anti-apoptotic effect of BclxL. D2 also reduced cell viability but less drastically than D1. Of note, no significant differences were found between the samples transfected with eGFP-D1 or eGFP-D2 or cells transfected with an empty plasmid when treated with doxorubicin, indicating that both D1 and D2 are capable of inhibiting BclxL function (Figure 7a). Western blot analysis confirmed comparable expression levels for the eGFP-T20, eGFP-D1, eGFP-D2, and eGFP-xL chimeras (Figure 7c). Furthermore, the fluorescence levels of all four eGFP chimeras were comparable (Figure 7d).

Figure 6. Subcellular localization of the designed inhibitors. **a**. The eGFP was fused to the TMD of BclxL, D1, D2, or the TMD of T20 to create the eGFP-xL, eGFP-D1, eGFP-D2, and eGFP-T20 chimeras, respectively. These chimeras (green) together with a construct bearing BclxL fuse to the mCherry fluorescent protein (mCherry-BclxL, red) were transfected into HeLa and 24 hours later analyzed by confocal microscopy (n=3). DAPI staining is shown in blue. The right column of

each panel shows the co-localization of both signals (yellow, visible only when the images are merged). **b**. Overview of the organelle differential centrifugation workflow. In the assay, a series of differential ultracentrifugation steps was used to separate organelles and subcellular compartments. The proteins in each fraction were quantified by SWATH-MS analysis. **c**. Abundance profiles of BclxL (blue line) and eGFP-D1 (black dotted line) across all fractions. A representative result is shown. **d**. Principal Component Analysis of the abundance profiles across fractions obtained by differential centrifugation. The plot shows the 1962 proteins identified by SWATH-MS. The percentage in each axis represents the amount of total variability that PC1 and PC2 can explain. Organelle markers are colored. eGFP-D1 and BclxL are shown in white with a black contour, and the rest of the proteins (other) are shown in gray. A representative plot is shown.
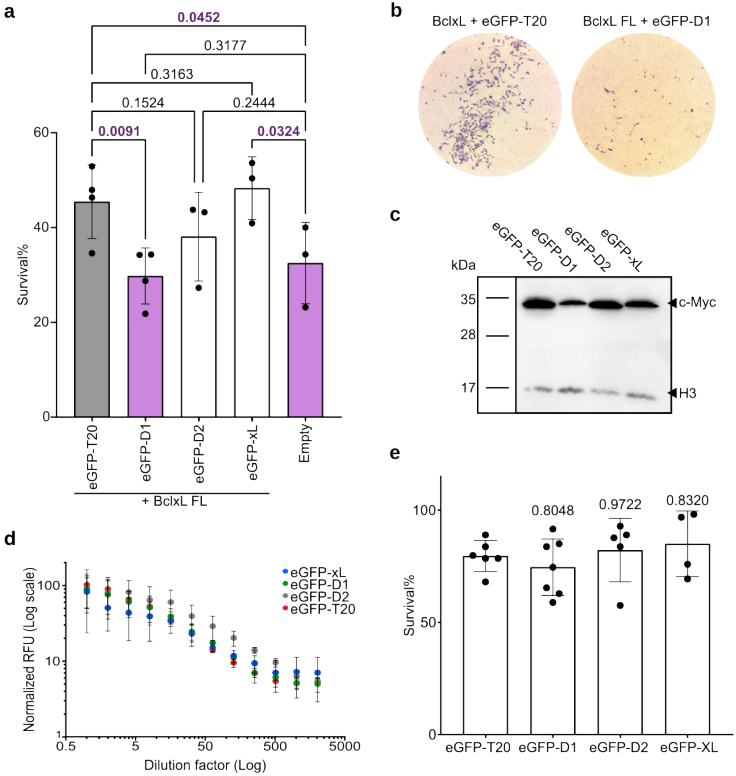
F**igure 7. Inhibition of the BclxL anti-apoptotic effect. a**. HeLa cells were transfected with BclxL together with either the eGFP-XL, eGFP-D1, eGFP-D2, or eGFP-T20 (gray bar, negative control) chimeras. Next, cells were treated with doxorubicin, and after 16 h the percentage of surviving cells was calculated based on Trypan blue staining. The survival percentage mean and standard deviation of at least three independent experiments are shown. Solid dots represent the results of

individual experiments. Transfection with an empty plasmid (Empty) was used as a negative control. Statistical differences are based on a one-tailed homoscedastic t-test (p-values are indicated above). Values significantly lower than those obtained with the T20 control are highlighted in purple. **b.** HeLa cells were co-transfected with Bclxl FL + eGFP-D1 or BclxL Fl + eGFP-D20 and treated with doxorubicin. After 16 h cells were washed, fixed, and stained. Images were taken using a 10x microscope objective. **c**. Western blot analysis of protein levels. Histone 3 (H3) was used as a loading control (n = 3). **d**. Fluorescence levels of the eGFP chimeras. HeLa cells were transfected with the indicated eGFP chimeras. After 24 hours samples were serially diluted and the relative fluorescence units (RFUs) and measured (λexc 485 nm, λem 535nm). The mean and standard deviation of at least three independent experiments are shown. **e**. HeLa cells were transfected for analysis of the toxicity associated with the expression of the eGFP-XL, eGFP-D1, eGFP-D2, or eGFP-T20 chimeras. The percentage of surviving cells was calculated based on Trypan blue staining 24 hours after transfection. The survival percentage mean and standard deviation of at least five independent experiments are shown. Solid dots represent the results of individual experiments. For statistical purposes, eGFP-T20 was used as a control. Statistical differences are based on a two-tailed homoscedastic t-test, p-values indicated above the corresponding bar.

The western blot analysis and the fluorescence levels of the eGFP-TMD chimeras suggested that none of them were toxic. Thus, the reduction in cell viability observed when they were transfected alongside BclxL and treated with doxorubicin was likely the result of their inhibition of BclxL and not a direct effect on cell viability. Nonetheless, we decided to test the viability of the cells after the transfection with eGFP-T20, eGFP-D1, eGFP-D2, or eGFP-xL. Our results indicated that neither D1 nor D2 was toxic to HeLa cells, a vital characteristic when designing a BclxL non-toxic inhibitor. (Figure 7e).

Discussion

Many methods, including computational approaches[67–69], have been developed to explore interactions among water-soluble proteins or regions of proteins. However, techniques to target TMDs have not been examined thoroughly because of the intrinsic difficulty of working with membrane proteins and lack of understanding about TMD–TMD interactions. In a seminal work, Yin et al. described a method for the computational design of peptides that target TMDs in a sequence-specific manner[18]. To illustrate their method, the authors designed and tested peptides that recognized the TMD of two integrins *in vitro* and *in vivo*. In more recent work, Elazar et al.

designed single-pass α-helical TMDs that self-assembled, basing their approach on an improved *ab initio* Rosetta atomistic modeling strategy[42,43,70]. These TMDs were incorporated into a chimeric antigen receptor (CAR), and results indicated that *in vitro* CAR T-cell cytokine release and *in vivo* anti-tumor activity scaled linearly with the oligomeric state encoded by the receptor TMD[19].

In this work, we take the identification and modulation of TMD–TMD interactions one step further. First, we identified the role of intramembrane interactions of BclxL during apoptosis and as potential therapeutical targets. Next, we designed short hydrophobic regions that insert into the membrane and selectively and strongly interact with the TMD of BclxL. Finally, we proved that the expression of our designs can indeed inhibit the anti-apoptotic function of BclxL.

Several methods have been described for the identification of TMD–TMD interactions, including *in vitro, in vivo,* and *in silico* approaches[71]. To determine whether the TMD of BclxL interacts with other TMDs from the Bcl2 family, we chose BLaTM[30] and an adapted BiFC assay[35]. The combination of these two methodologies allowed us to quantitatively analyze potential TMD–TMD interactions (BLaTM) and investigate their presence *in vivo* in eukaryotic cells (BiFC). Furthermore, the adapted BiFC assay replicated the native membrane topology of the proteins under study. Bcl2 proteins are characterized by the absence of a signal peptide and by a single TMD in their Ct end[53], this type of membrane proteins are known as tail-anchored proteins[56]. By locating the TMD at the Ct end of our BiFC chimeras, we replicated the location/orientation and membrane targeting of a Bcl2 protein and, we inferred, the necessary features for TMD–TMD interaction.

Both the BLaTM and BiFC assays facilitated the analysis of each potential TMD interaction independently. There are many advantages of reductionist and isolating approaches. Nevertheless, there are some limitations to consider, including that the formation of higher-order oligomers (e.g., trimers, tetramers) could not be excluded. Additionally, although the

formation of homo-oligomers is self-sufficient (no other TMD is involved), the formation of a hetero-oligomer might require prior assembly of an intramembrane homo-oligomer, and BiFC or BLaTM do not yield the relevant information on this matter.

The results obtained with the BiFC and BLaTM approaches were very similar, nonetheless, we observed some differences. These differences can largely be explained by discrepancies in the nature and location of reporters and, more important, in the membrane environment, prokaryotes vs. eukaryotes, which affects TMD–TMD interactions[35,72–74]. The disparity observed for the BclxL and Bax TMD-TMD interaction could also be attributed to the low expression levels of Bax TMD in 293T cells. Expression of Bax TMD domain has been shown to induce some toxicity in eukaryotic cells[12,14].

There have been described 3 modes in which BclxL could block apoptosis. Mode 0 suggests that BclxL shift the equilibrium between soluble and membrane-bound Bax[62,63]. Mode 1 indicates that BclxL sequesters BH3 activators[75,76]. Lastly, Mode 2, proposes that BclxL directly sequester pro-apoptotic Bax[76–78]. We detected interactions between the TMD of BclxL and the TMD of pro-apoptotic, anti-apoptotic, and BH3-only Bcl2 members. Furthermore, a membrane-bound BclxL chimera incapable of establishing intramembrane TMD-TMD interactions (such as Bclxl-T20) showed no anti-apoptotic properties. Of interest, the elimination of TMD–TMD interactions by substitution of the BclxL TMD with the TMD of mitochondrial Tomm20 protein had a stronger effect on BclxL inhibition of doxorubicin-induced apoptosis[38] than did the complete elimination of the TMD. Thus, while we have no data regarding Mode 0 for the induction of apoptosis, in membranes, both Mode 1 and 2 are compatible with our results. Furthermore, we provided data suggesting that in Mode 1 and/or 2 the interactions of the TMD of BclxL with other TMDs of the Bcl2 family members play a key role. Of note, *in vivo*, all the described TMD–TMD interactions could occur simultaneously or display a hierarchical order in which some interactions are preferred. Currently, data exist to support either possibility[15,79], and the two are not mutually exclusive and most likely occur simultaneously in cells.

To block the intramembrane interactions of BclxL and consequently its anti-apoptotic function, we designed three potential inhibitors (D1, D2, and D3), based on a model of a putative BclxL TMD homodimer. Our data does not inform on the oligomeric state of BclxL TMD, however, there is data suggesting that BclxL forms homo-dimers both in solution and in the membrane[15]. All three inhibitors' designs retained the Gly residues within the TMD of BclxL (Gly8 and Gly13, D1) or substituted one of them with an Ala residue (Gly8Ala, D2 and D3). Small residues, particularly Gly, have been observed in other TMD–TMD oligomers (e.g., GpA[31,32], integrins[30], and even viral Bcl2 proteins[14]). The presence of these small residues in the designs facilitates helix–helix packing, maximizing the surface area and thus van der Waals forces, a major contributor to both homo- and hetero-intramembrane PPIs[80].

According to the results of the BLaTM assay, D2 forms a weaker interaction with the TMD of BclxL than does D1. The only difference between these two designs is in position 8, where there is a Gly in D1 and an Ala in D2. Our results (for both the BlaTM and the functional assays) indicate that a small structural difference, the substitution of Gly by Ala, has a major impact on the TMD–TMD interaction and the associated biological processes. These results, together with the substitutions at position 12, imply a high specificity in intramembrane interactions. Note that high specificity is necessary to achieve a complex interaction network like the one suggested by our TMD–TMD PPI scanning. However, it remains to be seen whether a single interaction can be interrupted specifically, which would indicate that each TMD–TMD PPI has its unique interaction profile. Most likely, given the reduced surface area of a TMD and the potentially large number of interaction partners, as is the case for the TMD of BclxL, some residues might be involved in interaction with multiple TMDs.

The expression of eGFP-D1 could block the anti-apoptotic function of BclxL. Likewise, D2 reduced the anti-apoptotic response of BclxL, although we found no statistically significant differences with the negative control. These results indicate that the strength of the interaction between D1 or D2

with the TMD of BclxL correlates with the inhibiting properties of the designs.

The subcellular localization by differential centrifugation in tandem with SWATH-MS suggests an organelle bound and a cytosolic form for BclxL (and Bax, see Supp. Information). It has been previously shown that Bcl2 family members shuttle between the mitochondria membrane and a soluble state[62–64]. On the other hand, D1 behaves exclusively as organelle bound protein. Thus, the D1-BclxL interaction and the corresponding inhibitory effect must occur in mitochondrial membranes. Any anti-cancer therapy based on our designs will require a vector that facilitates the insertion of the inhibitor into the MOM. Given the peptide nature of the inhibitors, there are two main possibilities for their delivery into cancer cells and their mitochondria. One is to use delivery systems for peptide-based drugs[81], such as cell-penetrating peptides (e.g., TAT), targeting peptides (e.g., RGD-4c), or stimuli-responsive peptides (e.g., pHLIP). The other is that instead of the peptide being delivered, a nucleic acid encoding the inhibitor (e.g., RNA/DNA lipid nanoparticles, polyplexes, or a viral vector) could be delivered. The hydrophobic nature of our designs complicates conventional peptide delivery, leading us to believe that the second option would be more appropriate. Of note, we could demonstrate that when eGFP-D1 is expressed from a DNA plasmid, it displays the same localization as BclxL in eukaryotic cells. This result indicates that the designs inherently can reach their active location when translated within cells with no further assistance.

In summary, we have provided evidence of the importance of TMD–TMD interactions in apoptosis control, particularly in the case of BclxL. We successfully designed sequences capable of specifically sequestering the TMD of BclxL. With this sequestration capacity, our designs could inhibit the anti-apoptotic action of BclxL. Our work shows a path to design effective inhibitors based only on the sequences of the target receptor. The fact that two of three designs exhibited the desired hetero and no homo-interactions highlights the accuracy of the TMHOP modeling strategy and the FuncLib design algorithm, which has already been applied to a wide range of soluble protein design tasks. We hope that this work both advances

our understanding of sequence-specific recognition in membranes and opens the way to a new generation of anti-cancer drugs without the limitations of current BclxL inhibitors.

Materials and Methods

Cell cultures, plasmids, and reagents

Human embryonic kidney 293T cells (HEK 293T), and human epithelial cervical cancer cells (HeLa) were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), and penicillin-streptomycin (P/S) (100 U/mL) (Gibco). All cells were grown at 37 °C, with 5% CO2. Transfection of DNA into eukaryotic cells was performed in Opti-MEM reduced serum medium (Gibco) with Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications.

The TMD sequences were synthesized by Invitrogen (GeneArt gene synthesis), PCR amplified, and subcloned into the appropriated vector either using a standard digestion-ligation protocol or using the InFusion cloning system following the manufacturer's protocol (Takara). Mutations into the TMD were introduced by site-directed mutagenesis using the Quick Change II kit following the manufacturer's instructions (Agilent Technologies). All DNA manipulations were confirmed by the sequencing of plasmid DNAs (Macrogen Spain). Transfection of DNA into eukaryotic cells was performed in Opti-MEM reduced serum medium (Gibco) with Lipofectamine 2000 (Invitrogen) according to the manufacturer's specifications.

Bimolecular fluorescent complementation (BiFC) assay

For the generation of BiFC chimeric plasmids including the Nt or Ct of the Venus Fluorescent Protein (VN, VC, respectively) plasmids were modified (Addgene #27097, #22011, a gift from Chang-Deng H)[82] to clone the cellular and viral Bcl2 TMDs at the Ct of the VFP[14]. Chimeras (500 ng VN + 500 ng VC) were transfected into 2.10^5 HEK 293T cells together with a plasmid expressing Renilla luciferase under the CMV promoter (pRL-CMV) (50 ng) for signal normalization. Cells were incubated at 37 °C, 5% CO_2 for 48h, PBS washed, and collected for fluorescence and luciferase measurements

(Victor X3 plate reader). For the Renilla luciferase readings, we used the Renilla Luciferase Glow Assay Kit (Pierce, Thermofisher) according to the manufacturer's protocol. In each experiment, the fluorescence/luminescence ratio obtained with the GpA homodimer was used as a 100% oligomerization value and the rest of the values were adjusted accordingly. All experiments were done at least in triplicates.

BLaTM assay

Competent *E. coli* BL21-DE3 cells were co-transformed with N-BLa and C-BLa plasmids, version 1.1[30], containing a given TMD pair and grown overnight at 37 °C on LB-agar plates containing 34 µg/mL of chloramphenicol (Cm) and 35 µg/mL of kanamycin (Kan) for plasmid inheritance. After o/n incubation at 37 °C, colonies were either picked for immediate use or the plates were sealed with Parafilm (Pechiney Plastic Packaging) and stored at 4 °C for up to one week. Overnight cultures were conducted by inoculating 5 mL of LB-medium (Cm, Kan) with 10 colonies from one agar plate, followed by o/n incubation in an orbital incubator at 37 °C, 200 rpm. An expression culture was started with a 1:10 dilution of the overnight culture in 4 mL expression medium: LB-medium (Cm, Kan) containing 1.33 mM arabinose. After 4 h at 37 °C, the expression cultures were diluted to an $OD_{600} = 0.1$ in the expression medium. To expose the bacteria to different ampicillin concentrations, an LD_{50} culture was prepared by pipetting 100 µL of the diluted expression culture into each cavity of a 96-deep well plate (96 square well, 2 mL, VWR) containing 400 µL of expression media (final $OD_{600} = 0.02$). Freshly prepared ampicillin stock (100 mg/mL in ethanol) was added, resulting in ampicillin concentrations ranging from 0 to 350 µg/mL, depending on the affinity of the TMD under investigation. As a rule, the maximum ampicillin concentration to be used for a particular case should be about twice the mean LD_{50}. The plates were incubated in a moisturized container for 16 h at 37 °C and 250 rpm on a shaker (shaking amplitude 10 mm, KS 260 Basic, IKA) containing tips in every well to ensure proper agitation. Cell density was measured via absorbance at 544 nm in a microplate reader (Victor X3, Perkin Elmer). To minimize clonal variation, at least two transformations were done and at least two separate LD_{50} cultures were inoculated from each batch of

transformed bacteria using ten colonies for each culture. Thus, at least 40 colonies entered each determination of LD_{50}. To measure and collect LD_{50} values from the dose-response curves, we used Prism 9 from GraphPad.

To analyze the expression levels of the chimeras, competent *E. coli* BL21-DE3 cells were transformed with one N-BLa or C-BLa plasmid, version 1.1, containing a given TMD and grown overnight at 37 °C on LB-agar plates containing 34 µg/mL of Cm (for N-BLa) or 35 µg/mL of Kan (for C-BLa) for plasmid inheritance. After o/n incubation at 37 °C, cultures were conducted by inoculating 5 mL of LB-medium (Cm or Kan) with 10 colonies from one agar plate, followed by o/n incubation in an orbital incubator at 37 °C, 200 rpm. An expression culture was started with a 1:10 dilution from the overnight culture in 4 mL expression medium: LB-medium (Cm or Kan) containing 1.33 mM arabinose. After 4 h at 37 °C, the expression cultures were diluted to an OD600 = 0.1 in 5 ml of expression medium (final volume) and grown o/n, 37 °C, 200xg. The morning after, 100 µL from each culture were transferred to a black 96-well plate to measure the fluorescence (Victor X3, Perkin Elmer).

In vitro transcription and translation
The Lep-derived constructs were assayed using the TNT T7 Quick Coupled System (#L1170, Promega). Each reaction containing 1 µL of PCR product, 0.5 µL of EasyTag™ EXPRESS 35S Protein Labeling Mix (Perkin Elmer) (5.5 µCi), and 0.3 µL of microsomes (tRNA Probes) was incubated at 30 °C for 90 min. Samples were analyzed by SDS-PAGE. The bands were quantified using a Fuji FLA-3000 phosphoimager and the Image Reader 8.1 software. Free energy was calculated using: $\Delta Gapp = -RT.lnKapp$, where $Kapp = f2g/f1g$ being f1g and f2g the fraction of single glycosylated and double glycosylated protein, respectively. Endoglycosidase H treatment (Roche) was carried out according to the specifications of the manufacturer.

Cell-viability assays
To measure doxorubicin-induced apoptosis 1.5×10^6 HeLa cells were plated in a 24 wells plate containing 0.5 ml of media in each well. After overnight incubation, each well was transfected in triplicates with 500 ng of DNA.

After 24 h of expression, cells were treated with doxorubicin (stock 2 mM in DMSO) achieving a final concentration of 15 µM. Approximately, 16 h post-treatment cells (including those in the supernatant) were collected and their viability was measured using Trypan blue and using an automated cell counter (Invitrogen, Countess™ II). At least, 2 measurements per well were done. Alternatively, 16/24 h post-treatment media was removed and the cells washed (PBS x2) and fixed with 6% Glutaraldehyde containing 0.5% of crystal violet. After 20 min fixed cells were washed with water (x3) before visualization.

To measure the anti-BclxL effect of the designs, 1.5×10^6 HeLa cells were plated in a 24-well plate (0.5 ml of media in each well). After overnight incubation, cells were transfected (in triplicate) with 500 ng of either eGFP-XL, eGFP-T20, eGFP-D1, eGFP-D2, or empty pCAGGS (negative control). Additionally, cells were co-transfected with 500 ng of BclxL full length except for the negative control where 500 ng of empty pCAGGS were transfected. After 24 h of expression, cells were treated with doxorubicin (stock 2 mM in DMSO) achieving a final concentration of 15 µM. Approximately, 16 h post-treatment cells (including those in the supernatant) were collected and their viability was measured using Trypan blue and using an automated cell counter (Invitrogen, Countess™ II). At least, 2 measurements per well were done.

To study the toxicity of individual chimeras, 1.5×10^6 HeLa cells were plated in a 24 wells plate containing 0.5 ml of media in each well. After overnight incubation, each well was transfected in triplicates with 500 ng of a plasmid containing eGFP-T20, eGFP-D1, eGFP-D2, or eGFP-xL. 24 h post-transfection, cells were collected (including those in the supernatant) and their viability was measured using Trypan blue and using an automated cell counter (Invitrogen, Countess™ II). At least, 2 measurements per well were done.

To analyze the expression levels of the chimeras, 1.5×10^6 HeLa cells were plated in a 24 wells plate containing 0.5 ml of media in each well. After overnight incubation, each well was transfected in duplicates with 500 ng of

a plasmid containing eGFP-T20, eGFP-D1, eGFP-D2, or eGFP-xL. Cells were incubated at 37 °C, 5% CO2 for 24 h, PBS washed, and collected for fluorescence measurements (Victor X3 plate reader) or western blot analysis.

Inhibitor design
The design of the inhibitors was started with the modeling of the BclxL – BclxL homo-interaction. The modeling was done using TMHOP (Trans-membrane Homo Oligomer Predictor; https://TMHOP.weizmann.ac.il)[42]. Lowest energy models were selected and visually inspected. Using the selected model as the starting template, the design step was prepared using FuncLib[44] with the membrane filter by selecting mutations that obeyed the following rules: 1. Positive design for a heterodimer (Non-symmetric FuncLib; $\Delta\Delta G < +1$ Rosetta energy units; R.e.u.); 2. Negative selection for a homodimer (symmetric FuncLib; $\Delta\Delta G > +5$ R.e.u.). The resulting TMDs were manually selected after visualizing the predicted structures.

TMD interaction surface area and TMD-TMD crossing angle calculation
Amino acid residues in the interface were selected using InterfaceResidues.py script by Jason Vertrees (pymolwiki.org/index.php/InterfaceResidues). Then, surface area in square Angstroms was calculated using get_area command in The PyMOL (TM) Molecular Graphics System, Version 2.3.0 Schrodinger, LLC for the complete TMD and the selected interface area. The crossing angle between helices was calculated using AngleBetweenHelices.py script by Thomas Holder (pymolwiki.org/index.php/AngleBetweenHelices).

Subcellular protein localization by differential ultracentrifugation
The proteins of interest were expressed in the HEK-293T human cell line (37 ºC and 5% CO_2 with DMEM supplemented with FBS, Penstrep, and Amphotericin B). For each construct, two plates with 7×10^6 cells each were seeded. After 24 hours cells were transfected with plasmids encoding the BclXL or eGFP-D1 using PEI. After 24 hours cells were harvested, washed with PBS buffer and collected by trypsinization. Next, cells were centrifugated (200g 5min) and resuspended with 1.5 mL of lysis buffer (0.25 M sucrose, 10 mM HEPES pH 7.4, 2 mM EDTA, 2 mM magnesium acetate,

cOmplete protease inhibitor (Roche)). Next cells were lysed in an Isobiotech homogenizer using the 12 µm clearance size ball. Cells were passed through the homogenizer until 90% of cell disruption was achieved followed by centrifugation 2x200g 5min to eliminate intact cells. The supernatant was collected and a series of differential centrifugations were carried out 4ºC. Precisely, 1.000 g 10 min, 3.000g 10 min, 5.000 g 10 min, 9,000 g 15 min, 12.000 g 15 min, 15.000 g 15 min, 30.000 g 20 min, 79.000 g 43 min, 120.000 g 45 min. Pellets were resuspended (8 M urea, 0.15% SDS, 50 mM HEPES pH 8.5) and the protein concentration was determined by Bradford. All samples were submitted for SWATH-MS quantification at the Proteomics core facility of the University of Valencia. Briefly, samples were diggested with sequencing grade trypsin (Promega) at 37º C o/n. The trypsin digestion was stopped with 10% TFA and the peptide mixtures were dried in a speed vacuum and re-suspended with 2% ACN; 0.1% TFA at a final concentration of 0.5 µg/µL. Next, 3 µl of peptide mixture sample were loaded onto a trap column (3µ C18-CL, 350 µm x 0.5 mm; Eksigent) and desalted with 0.1% TFA at 5 µl/min during 5 min. The peptides were then loaded onto an analytical column (3µ C18-CL 120 Ǎ, 0.075 x 150 mm; Eksigent) equilibrated in 5% acetonitrile 0.1% FA (formic acid). Elution was carried out with a linear gradient of 7-40% B in A for 30 min for in gel digested samples or 120 min for in solution digested samples. (A: 0.1% FA; B: ACN, 0.1% FA) at a flow rate of 300 nL/min. Peptides were analyzed in a mass spectrometer nanoESI qQTOF (6600plus TripleTOF, ABSCIEX). Samples were ionized in a Source Type: Optiflow < 1 uL Nano applying 3.0 kV to the spray emitter at 175 ºC. Analysis was carried out in a data-dependent mode. Survey MS1 scans were acquired from 350–1400 m/z for 250 ms. The quadrupole resolution was set to 'LOW' for MS2 experiments, which were acquired 100–1500 m/z for 25 ms in 'high sensitivity' mode. Following switch criteria were used: charge: 2+ to 4+; minimum intensity; 250 counts per second (cps). Up to 100 ions were selected for fragmentation after each survey scan. Dynamic exclusion was set to 15 s. The rolling collision energies equations were set for all ions as +2 ions, according to the following equations: $|CE|=(slope)x(m/z)+(intercept)$.

ProteinPilot default parameters were used to generate peak list directly from 6600 plus TripleTOF wiff files. The Paragon algorithm of ProteinPilot v 5.0 was used to search the SwissProt database (564631 proteins) with the following parameters: trypsin specificity, IAM cys-alkylation, taxonomy restricted to homo sapiens and the search effort set to rapid with FDR analysis. The protein grouping was done by Pro group algorithm (a protein group in a Pro Group Report is a set of proteins that share some physical evidence. Unlike sequence alignment analyses where full-length theoretical sequences are compared, the formation of protein groups in Pro Group is guided entirely by observed peptides only. Since the observed peptides are determined from experimentally acquired spectra, the grouping can be considered to be guided by usage of spectra. Then, unobserved regions of protein sequence play no role in explaining the data).

Data analysis was performed using the R Bioconuctor[60,83] packages MSnbase v2.6.1[84] and pRoloc v1.21.9[85] as described previously[60]. Briefly, 334 marker proteins from these packages were found in our samples and were used to define 9 subcellular locations: cytosol, proteasome, nucleus, ribosome 40S, ribosome 60S, mitochondrion, plasma membrane, and endoplasmic reticulum. Subcellular locations that were defined by less than 13 markers were discarded from the analysis. To predict the subcellular localization of the unlabelled proteins we performed a classification using a Support Vector Machine (SVM) with a radial basis function kernel included in the package. Using the R code[61] we performed 100 rounds of fivefold cross-validation to estimate the performance of the algorithm. The optimization of the sigma and cost-free parameters was determined using the pair of sigma/cost provided by the biggest F1 score (harmonic mean of precision and recall). The proteins are assigned to a subcellular location based on the SVM score. To reduce the number of false positives we set a third quantile organelle score as a threshold[60].

Confocal microscopy
Confocal micrographs were done at the Microscopy Core Facility of the SCSIE (University of Valencia) using an Olympus FV1000 confocal microscope with a ×60 oil lens. Mitochondria mCherry fluorescent-labeled

marker (mCherry-Mito) was obtained from Addgene plasmid repository #55146, a gift from Michael Davidson, Institute of Molecular Biophysics and Center for Materials Research and Technology, The Florida State University. HeLa cells (5×10^3 cells/well) were seeded on 10 mm cover-slides treated with poly-Lys and placed in 24-well plates. The next day, cells were transfected with the appropriate plasmids. After 24 h, the cells were fixed (4% paraformaldehyde) and DAPI stained before image capture. A 1:1000 dilution in TBS 0.005% Tween Rabbit anti-c-Myc (Sigma PLA0001) antibody followed by an anti-Rabbit Alexa 488 conjugated (Life Technologies A21206) (1:1000) was used to label BclxL FL, BclxL T20, and BclxL ΔTM proteins. For the eGFP chimeras: eGFP-xL, eGFP-D1, eGFP-D1, and eGFP-D2 no antibody was needed as it was possible to use the eGFP itself. Pictures were taken in an Olympus FV1000 confocal microscope. Laser intensity was individually adjusted in all samples. Pictures were not used for quantification.

Western blot analysis
Cell monolayers were lysed in Laemmli's sample buffer (62.5 mM Tris-HCl [pH 6.8], 2% sodium dodecyl sulfate [SDS], 0.01% bromophenol blue, 10% glycerol and 5% β-mercaptoethanol). Protein samples were subjected to 12 % SDS-polyacrylamide gel electrophoresis (PAGE) and transferred to nitrocellulose membranes (BioRad). Membranes were blocked for 30 min at room temperature in Tris-buffered saline supplemented with 0.05% Tween 20 (TBS-T) containing 5% non-fat dry milk and later incubated with primary antibodies diluted in the same buffer at 4 °C overnight. Antibodies used in this study were β-actin (Santa Cruz Biotechnology SC-47778), c-Myc (Sigma PLA0001 or Roche 11667149001), Histone 3 (Sigma H0164), and Flag (Sigma B3111). Then, membranes were washed with TBS-T and incubated with goat anti-mouse IgG-peroxidase conjugate (Sigma DC02L) for 1 h at room temperature and washed again. All antibodies were used at a 1:10,000 dilution in TBS-T with 5% non-fat dry milk. Detection of immunoreactive proteins was carried out using the enhanced chemiluminescence (ECL) reaction (SuperSignal ThermoScientific) and detected by the ChemiDoc Touch Imaging System (BioRad).

Competing interest

The University of Valencia has filed a patent application related to the use of the described sequences for the treatment of cancer.

References

1. Hengartner, M. O. The biochemistry of apoptosis. Nature 407, 770–776 (2000).
2. Alnemri, E. S. et al. Human ICE/CED-3 Protease Nomenclature. Cell 87, 171 (1996).
3. Fulda, S. & Debatin, K.-M. Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. Oncogene 25, 4798–4811 (2006).
4. Kim, H. et al. Hierarchical regulation of mitochondrion-dependent apoptosis by BCL-2 subfamilies. Nat. Cell Biol. 8, 1348–1358 (2006).
5. Hafezi, S. & Rahmani, M. Targeting BCL-2 in Cancer: Advances, Challenges, and Perspectives. Cancers 13, 1292 (2021).
6. Delbridge, A. R. D., Grabow, S., Strasser, A. & Vaux, D. L. Thirty years of BCL-2: translating cell death discoveries into novel cancer therapies. Nat. Rev. Cancer 16, 99–109 (2016).
7. Oltvai, Z. N., Milliman, C. L. & Korsmeyer, S. J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. Cell 74, 609–619 (1993).
8. Wang, K., Yin, X. M., Chao, D. T., Milliman, C. L. & Korsmeyer, S. J. BID: a novel BH3 domain-only death agonist. Genes Dev. 10, 2859–2869 (1996).

9.   Kelekar, A., Chang, B. S., Harlan, J. E., Fesik, S. W. & Thompson, C. B. Bad is a BH3 domain-containing protein that forms an inactivating dimer with Bcl-XL. Mol. Cell. Biol. 17, 7040–7046 (1997).

10.   Cosentino, K. & García-Sáez, A. J. Bax and Bak Pores: Are We Closing the Circle? Trends Cell Biol. 27, 266–275 (2017).

11.   Dadsena, S., King, L. E. & García-Sáez, A. J. Apoptosis regulation at the mitochondria membrane level. Biochim. Biophys. Acta BBA - Biomembr. 1863, 183716 (2021).

12.   Andreu-Fernández, V. et al. Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. Proc. Natl. Acad. Sci. U. S. A. 114, 310–315 (2017).

13.   Lucendo, E. et al. Mcl-1 and Bok transmembrane domains: Unexpected players in the modulation of apoptosis. Proc. Natl. Acad. Sci. U. S. A. 117, 27980–27988 (2020).

14.   García-Murria, M. J. et al. Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control cellular apoptosis. Nat. Commun. 11, 6056 (2020).

15.   Bleicken, S., Hantusch, A., Das, K. K., Frickey, T. & Garcia-Saez, A. J. Quantitative interactome of a membrane Bcl-2 network identifies a hierarchy of complexes for apoptosis regulation. Nat. Commun. 8, 73 (2017).

16.   Bleicken, S. et al. Structural model of active Bax at the membrane. Mol. Cell 56, 496–505 (2014).

17.   Zhang, Z. et al. BH3-in-groove dimerization initiates and helix 9 dimerization expands Bax pore assembly in membranes. EMBO J. 35, 208–236 (2016).

18.   Yin, H. et al. Computational design of peptides that target transmembrane helices. Science 315, 1817–1822 (2007).

19.   Elazar, A. et al. De novo-designed transmembrane domains tune engineered receptor functions. eLife 11, e75660 (2022).

20.   Kaufmann, S. H. et al. Elevated Expression of the Apoptotic Regulator Mcl-1 at the Time of Leukemic Relapse. Blood 91, 991–1000 (1998).

21.   Reed, J. C. Regulation of apoptosis by bcl-2 family proteins and its role in cancer and chemoresistance. Curr. Opin. Oncol. 7, 541–546 (1995).

22.		Lotem, J. & Sachs, L. Control of apoptosis in hematopoiesis and leukemia by cytokines, tumor suppressor and oncogenes. Leukemia 10, 925–931 (1996).

23.		Mesner, P. W., Budihardjo, I. I. & Kaufmann, S. H. Chemotherapy-induced apoptosis. Adv. Pharmacol. San Diego Calif 41, 461–499 (1997).

24.		Bhola, P. D. & Letai, A. Mitochondria-Judges and Executioners of Cell Death Sentences. Mol. Cell 61, 695–704 (2016).

25.		Lucianò, A. M., Pérez-Oliva, A. B., Mulero, V. & Del Bufalo, D. Bcl-xL: A Focus on Melanoma Pathobiology. Int. J. Mol. Sci. 22, 2777 (2021).

26.		Parry, N., Wheadon, H. & Copland, M. The application of BH3 mimetics in myeloid leukemias. Cell Death Dis. 12, (2021).

27.		Lessene, G., Czabotar, P. E. & Colman, P. M. BCL-2 family antagonists for cancer therapy. Nat. Rev. Drug Discov. 7, 989–1000 (2008).

28.		Souers, A. J. et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat. Med. 19, 202–208 (2013).

29.		Schoenwaelder, S. M. et al. Bcl-xL-inhibitory BH3 mimetics can induce a transient thrombocytopathy that undermines the hemostatic function of platelets. Blood 118, 1663–1674 (2011).

30.		Schanzenbach, C., Schmidt, F. C., Breckner, P., Teese, M. G. & Langosch, D. Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions. Sci. Rep. 7, 43476 (2017).

31.		Lemmon, M. A., Flanagan, J. M., Treutlein, H. R., Zhang, J. & Engelman, D. M. Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry 31, 12719–12725 (1992).

32.		MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. A transmembrane helix dimer: structure and implications. Science 276, 131–133 (1997).

33.		Orzáez, M., Pérez-Payá, E. & Mingarro, I. Influence of the C-terminus of the glycophorin A transmembrane fragment on the dimerization process. Protein Sci. Publ. Protein Soc. 9, 1246–1253 (2000).

34.   Kerppola, T. K. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. Nat. Protoc. 1, 1278–1286 (2006).

35.   Grau, B. et al. The role of hydrophobic matching on transmembrane helix packing in cells. Cell Stress 1, 90–106 (2017).

36.   Rooswinkel, R. W. et al. Antiapoptotic potency of Bcl-2 proteins primarily relies on their stability, not binding selectivity. Blood 123, 2806–2815 (2014).

37.   González-García, M. et al. bcl-XL is the major bcl-x mRNA form expressed during murine development and its product localizes to mitochondria. Dev. Camb. Engl. 120, 3033–3042 (1994).

38.   Fang, W., Rivard, J. J., Mueller, D. L. & Behrens, T. W. Cloning and molecular characterization of mouse bcl-x in B and T lymphocytes. J. Immunol. 153, 4388–4398 (1994).

39.   Zamzami, N., Brenner, C., Marzo, I., Susin, S. A. & Kroemer, G. Subcellular and submitochondrial mode of action of Bcl-2-like oncoproteins. Oncogene 16, 2265–2282 (1998).

40.   Martínez-Gil, L., Saurí, A., Marti-Renom, M. A. & Mingarro, I. Membrane protein integration into the endoplasmic reticulum. FEBS J. 278, 3846–3858 (2011).

41.   Chio, U. S., Cho, H. & Shan, S.-O. Mechanisms of Tail-Anchored Membrane Protein Targeting and Insertion. Annu. Rev. Cell Dev. Biol. 33, 417–438 (2017).

42.   Weinstein, J. Y., Elazar, A. & Fleishman, S. J. A lipophilicity-based energy function for membrane-protein modelling and design. PLoS Comput. Biol. 15, (2019).

43.   Elazar, A. et al. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. eLife 5, e12125 (2016).

44.   Khersonsky, O. et al. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. Mol. Cell 72, 178-186.e5 (2018).

45.   Netzer, R. et al. Ultrahigh specificity in a network of computationally designed protein-interaction pairs. Nat. Commun. 9, 5286 (2018).

46.   Bolenz, C. et al. Optimizing chemotherapy for transitional cell carcinoma by application of bcl-2 and bcl-xL antisense oligodeoxynucleotides. Urol. Oncol. 25, 476–482 (2007).

47.   Martínez-Gil, L. et al. Membrane integration of poliovirus 2B viroporin. J. Virol. 85, 11315–11324 (2011).

48.   Bañó-Polo, M., Martínez-Garay, C. A., Grau, B., Martínez-Gil, L. & Mingarro, I. Membrane insertion and topology of the translocon-associated protein (TRAP) gamma subunit. Biochim. Biophys. Acta BBA - Biomembr. 1859, 903–909 (2017).

49.   Lundin, C., Kim, H., Nilsson, I., White, S. H. & von Heijne, G. Molecular code for protein insertion in the endoplasmic reticulum membrane is similar for N(in)-C(out) and N(out)-C(in) transmembrane helices. Proc. Natl. Acad. Sci. U. S. A. 105, 15702–15707 (2008).

50.   Kasturi, L., Chen, H. & Shakin-Eshleman, S. H. Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. 323 (Pt 2), 415–419 (1997).

51.   Mingarro, I., von Heijne, G. & Whitley, P. Membrane-protein engineering. Trends Biotechnol. 15, 432–437 (1997).

52.   Braunger, K. et al. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. Science 360, 215–219 (2018).

53.   Andreu-Fernández, V. et al. The C-terminal Domains of Apoptotic BH3-only Proteins Mediate Their Insertion into Distinct Biological Membranes. J. Biol. Chem. 291, 25207–25216 (2016).

54.   Hessa, T. et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450, 1026–1030 (2007).

55.   Hessa, T. et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433, 377–381 (2005).

56.   Hegde, R. S. & Keenan, R. J. The mechanisms of integral membrane protein biogenesis. Nat. Rev. Mol. Cell Biol. 23, 107–124 (2022).

57.   Geladaki, A. et al. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. Nat. Commun. 10, 331 (2019).

58.	Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. Proteomics 20, e1900276 (2020).

59.	Rotello, R. J. & Veenstra, T. D. Mass Spectrometry Techniques: Principles and Practices for Quantitative Proteomics. Curr. Protein Pept. Sci. 22, 121–133 (2021).

60.	Breckels, L. M., Mulvey, C. M., Lilley, K. S. & Gatto, L. A Bioconductor workflow for processing and analysing spatial proteomics data. F1000Research 5, 2926 (2016).

61.	Crook, O. M., Mulvey, C. M., Kirk, P. D. W., Lilley, K. S. & Gatto, L. A Bayesian mixture modelling approach for spatial proteomics. PLoS Comput. Biol. 14, e1006516 (2018).

62.	Schellenberg, B. et al. Bax exists in a dynamic equilibrium between the cytosol and mitochondria to control apoptotic priming. Mol. Cell 49, 959–971 (2013).

63.	Edlich, F. et al. Bcl-x(L) retrotranslocates Bax from the mitochondria into the cytosol. Cell 145, 104–116 (2011).

64.	Eskes, R., Desagher, S., Antonsson, B. & Martinou, J. C. Bid induces the oligomerization and insertion of Bax into the outer mitochondrial membrane. Mol. Cell. Biol. 20, 929–935 (2000).

65.	Thul, P. J. et al. A subcellular map of the human proteome. Science 356, eaal3321 (2017).

66.	Christoforou, A. et al. A draft map of the mouse pluripotent stem cell spatial proteome. Nat. Commun. 7, 9992 (2016).

67.	Cao, L. et al. Design of protein binding proteins from target structure alone. Nature (2022) doi:10.1038/s41586-022-04654-9.

68.	Bouchiba, Y., Ruffini, M., Schiex, T. & Barbe, S. Computational Design of Miniprotein Binders. Methods Mol. Biol. Clifton NJ 2405, 361–382 (2022).

69.	Delaunay, M. & Ha-Duong, T. Computational Tools and Strategies to Develop Peptide-Based Inhibitors of Protein-Protein Interactions. Methods Mol. Biol. Clifton NJ 2405, 205–230 (2022).

70.	Elazar, A., Weinstein, J. J., Prilusky, J. & Fleishman, S. J. Interplay between hydrophobicity and the positive-inside rule in determining

membrane-protein topology. Proc. Natl. Acad. Sci. U. S. A. 113, 10340–10345 (2016).

71.   Duart, G., Grau, B., Mingarro, I. & Martinez-Gil, L. Methodological approaches for the analysis of transmembrane domain interactions: A systematic review. Biochim. Biophys. Acta Biomembr. 1863, 183712 (2021).

72.   Petrache, H. I., Grossfield, A., MacKenzie, K. R., Engelman, D. M. & Woolf, T. B. Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. J. Mol. Biol. 302, 727–746 (2000).

73.   Flinner, N., Mirus, O. & Schleiff, E. The influence of fatty acids on the GpA dimer interface by coarse-grained molecular dynamics simulation. Int. J. Mol. Sci. 15, 14247–14268 (2014).

74.   Janosi, L., Prakash, A. & Doxastakis, M. Lipid-modulated sequence-specific association of glycophorin A in membranes. Biophys. J. 99, 284–292 (2010).

75.   García-Sáez, A. J., Ries, J., Orzáez, M., Pérez-Payà, E. & Schwille, P. Membrane promotes tBID interaction with BCL(XL). Nat. Struct. Mol. Biol. 16, 1178–1185 (2009).

76.   Llambi, F. et al. A unified model of mammalian BCL-2 protein family interactions at the mitochondria. Mol. Cell 44, 517–531 (2011).

77.   Hsu, Y. T. & Youle, R. J. Nonionic detergents induce dimerization among members of the Bcl-2 family. J. Biol. Chem. 272, 13829–13834 (1997).

78.   Cheng, E. H., Levine, B., Boise, L. H., Thompson, C. B. & Hardwick, J. M. Bax-independent inhibition of apoptosis by Bcl-XL. Nature 379, 554–556 (1996).

79.   Fitzsimmons, L. et al. EBV BCL-2 homologue BHRF1 drives chemoresistance and lymphomagenesis by inhibiting multiple cellular pro-apoptotic proteins. Cell Death Differ. 27, 1554–1568 (2020).

80.   Martinez-Gil, L. & Mingarro, I. Viroporins, Examples of the Two-Stage Membrane Protein Folding Model. Viruses 7, 3462–3482 (2015).

81.   Berillo, D., Yeskendir, A., Zharkinbekov, Z., Raziyeva, K. & Saparov, A. Peptide-Based Drug Delivery Systems. Medicina (Mex.) 57, 1209 (2021).

82.   Kodama, Y. & Hu, C.-D. An improved bimolecular fluorescence complementation assay with a high signal-to-noise ratio. BioTechniques 49, 793–805 (2010).

83.   Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, R80 (2004).

84.   Gatto, L. & Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinforma. Oxf. Engl. 28, 288–289 (2012).

85.   Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T. & Lilley, K. S. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. Bioinforma. Oxf. Engl. 30, 1322–1324 (2014).
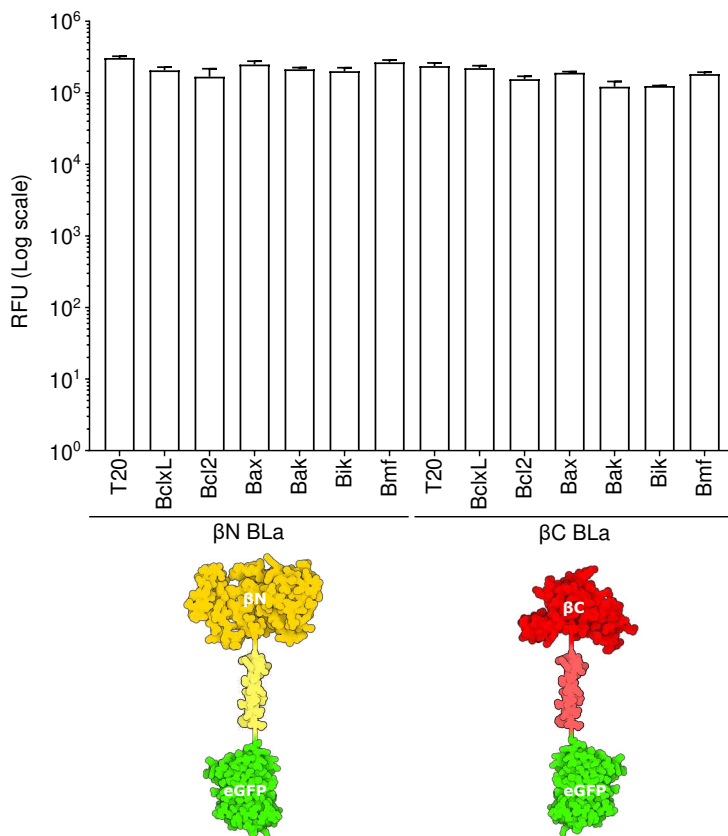
Figure S1. Analysis of the expression levels of the BLaTM chimeras. The eGFP-associated relative fluorescence (λ_{exc} 485 nm, λ_{em} 535nm) values for all BLaTM chimeras are shown. The mean and standard deviation of three independent experiments are shown (n = 3).
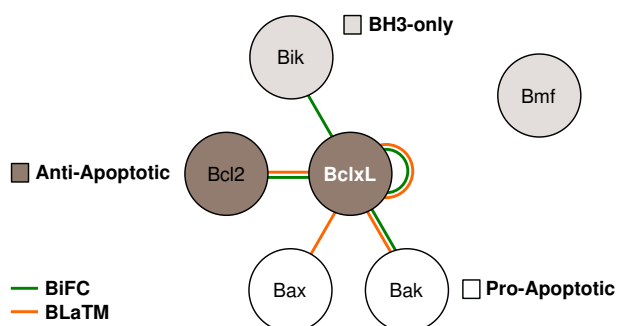
Figure S2. TMD-TMD interaction network of BclxL. A network representation of BclxL TMD interactions. The figure includes the results of the BiFC (green lines) and BLaTM assays (orange lines). Solid lines represent interactions, while TMDs are represented by nodes. Anti-apoptotic, pro-apoptotic, and BH3-only nodes are colored distinctively.
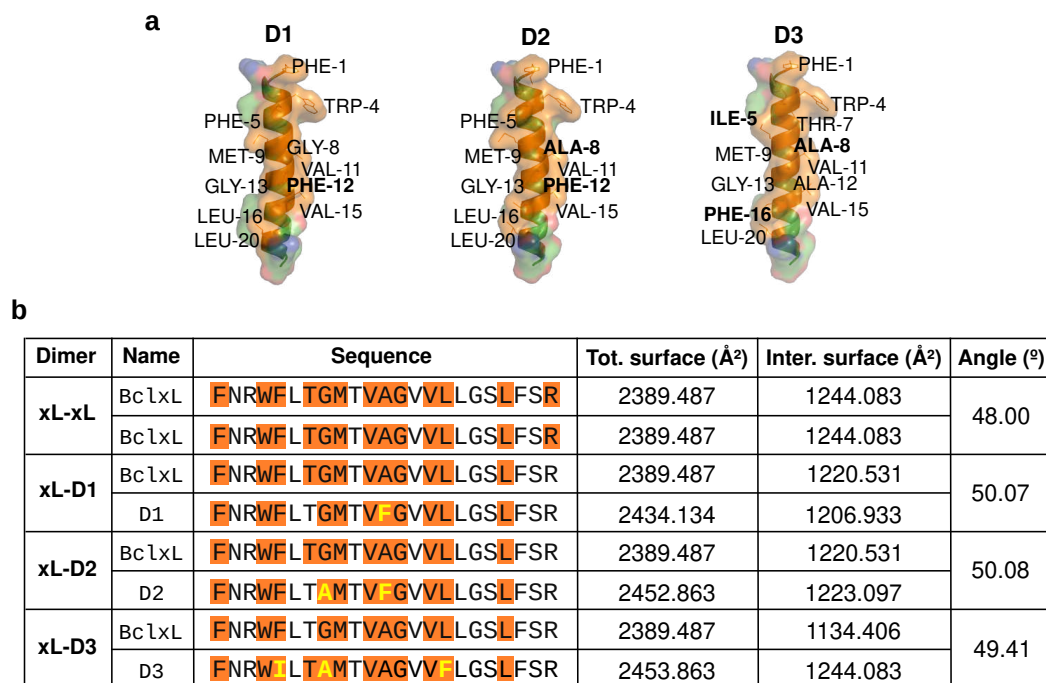
Figure S3. Interaction between the TMD of BclxL and D1, D2, and D3 designs. **a**. Structural representations of D1, D2, and D3 designs. Numbers correspond to the position of the amino acids in the design. Differences among the designs are highlighted in bold. **b**. The amino acids on the interaction surfaces of the potential homo and heterodimers are marked in orange. Changes in the designs over the TMD sequence of BclxL are highlighted in yellow. The total surface area of each monomer and the area buried in the interaction with the corresponding TMD in a potential dimer are shown (Å²). Additionally, the crossing angle of the monomers for each interaction is indicated.
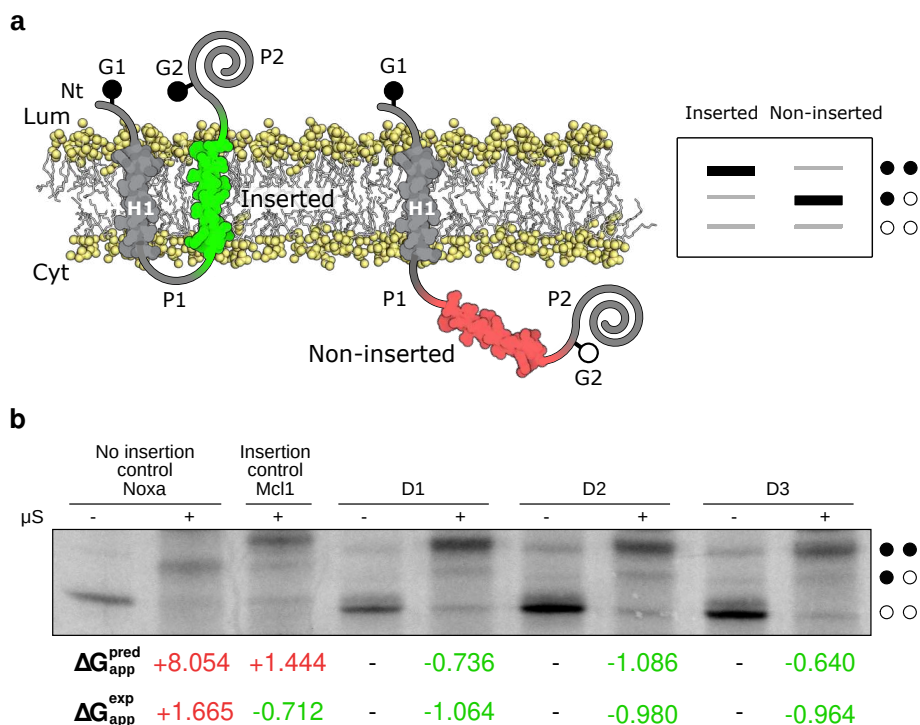
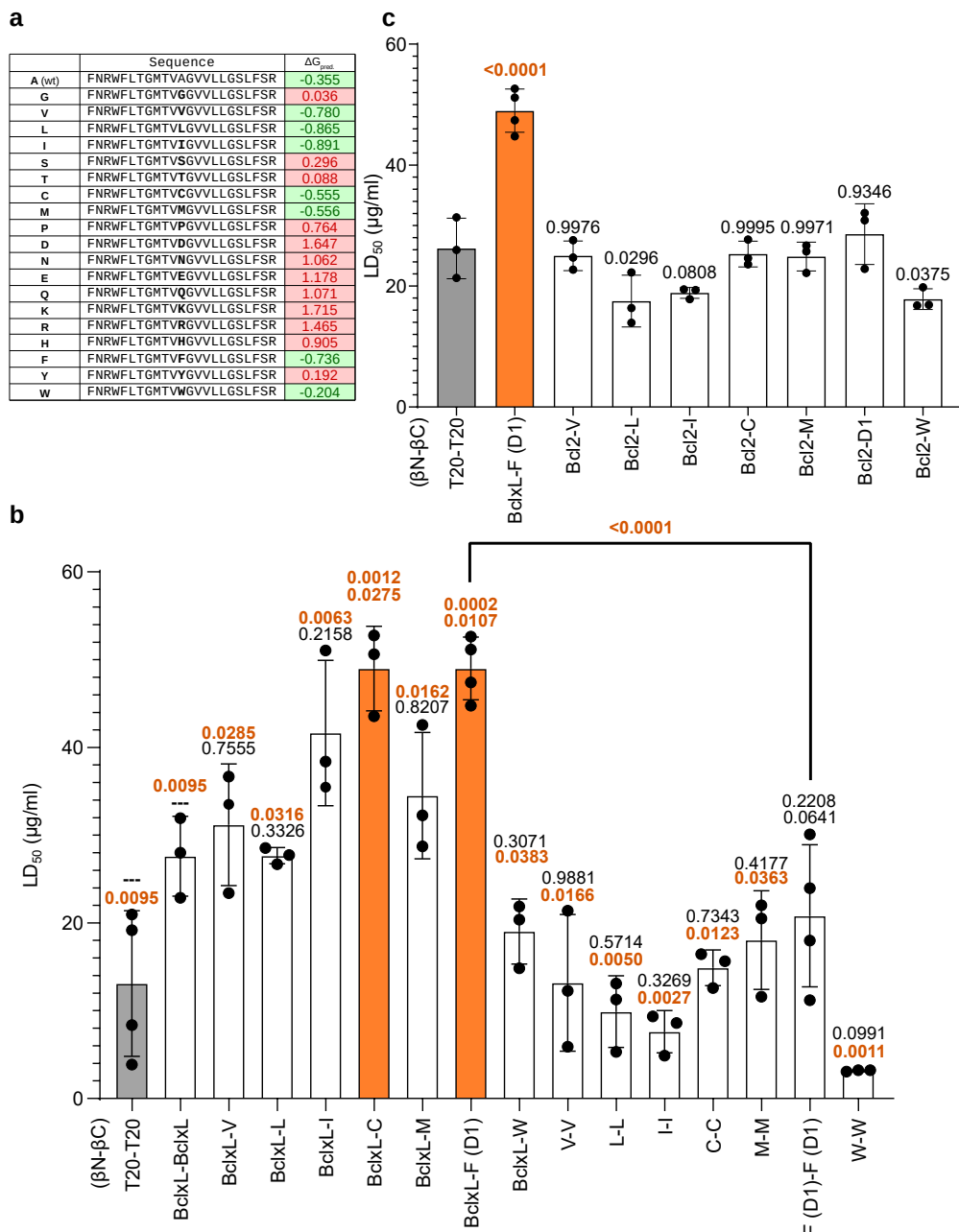**Figure S4. Insertion of D1, D2, and D3 in eukaryotic membranes. a**. Schematic representation of the leader peptidase (Lep) model protein. G1 and G2 denote artificial glycosylation acceptor sites. The sequence under investigation is introduced between the P1 and P2 domains (replacing the natural occuring H2 domain, not shown) of Lep. Recognition of the tested sequence by the translocon machinery as a TMD (highlighted in green) results in the modification of the G1 and G2 acceptor sites. In contrast, only G1 will be glycosylated if the sequence being tested is not recognized as a TMD (shown in red) and thus not inserted into ER-derived membranes. The glycosylation state of a protein can be monitored by SDS-PAGE based on the increase in molecular weight associated with the addition of a sugar moiety. A mock SDS-PAGE is represented on the right. The absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots. **b**. A representative example (n = 3) of in vitro protein translation in the presence (+) or absence (−) of ER-derived microsomes (µS). The absence of glycosylation of G1 and G2 acceptor sites is indicated by two white dots, single glycosylation by one white and one black dot, and double glycosylation by two black dots. Additionally, the predicted and experimental ΔGs for the insertion in kcal/mol (ΔG$_{app}^{pred}$ and ΔG$_{app}^{exp}$) are presented below the image. Experimental results are the average of at least three independent experiments. Green numbers indicate negative ΔGs (insertion of the tested sequence), and red numbers indicate positive ΔGs (no insertion of the tested sequence).

171
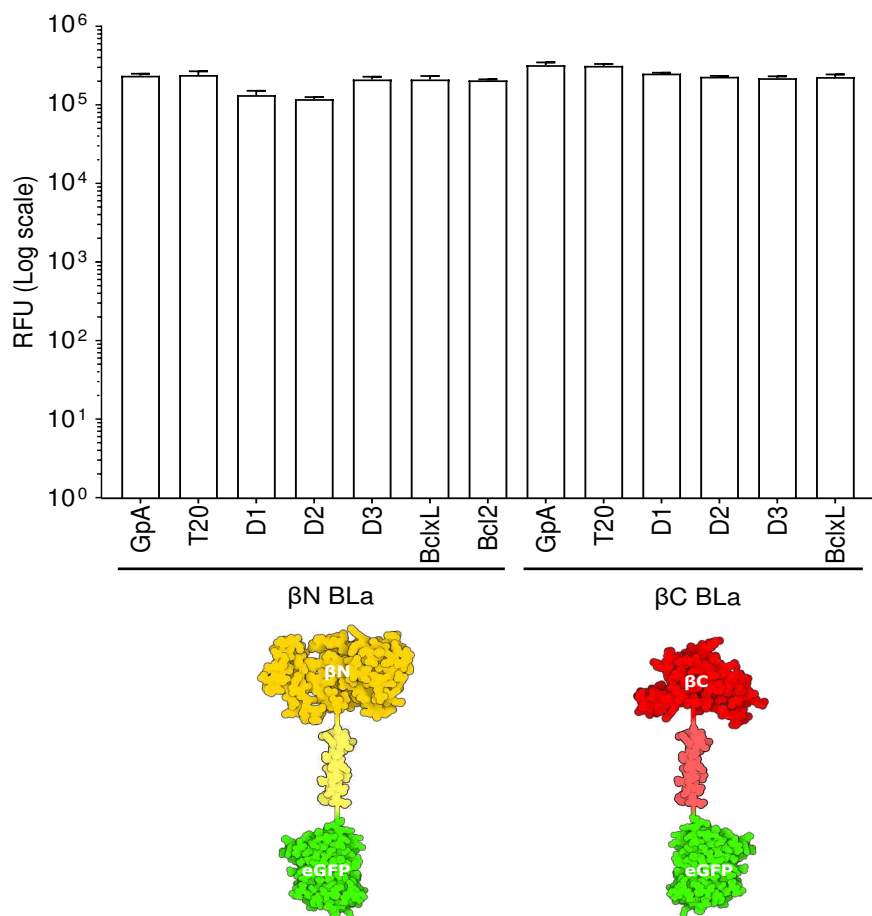
Figure S5. Analysis of TMD—TMD interactions by BLaTM. The eGFP-relative fluorescence (λ_{exc} 485 nm, λ_{em} 535nm) values for the BLaTM chimeras are shown. The mean and standard deviation of three independent experiments are shown (n = 3).

a

|	Sequence	$\Delta G_{pred.}$
A (wt)	FNRWFLTGMTV**A**GVVLLGSLFSR	-0.355
G	FNRWFLTGMTV**G**GVVLLGSLFSR	0.036
V	FNRWFLTGMTV**V**GVVLLGSLFSR	-0.780
L	FNRWFLTGMTV**L**GVVLLGSLFSR	-0.865
I	FNRWFLTGMTV**I**GVVLLGSLFSR	-0.891
S	FNRWFLTGMTV**S**GVVLLGSLFSR	0.296
T	FNRWFLTGMTV**T**GVVLLGSLFSR	0.088
C	FNRWFLTGMTV**C**GVVLLGSLFSR	-0.555
M	FNRWFLTGMTV**M**GVVLLGSLFSR	-0.556
P	FNRWFLTGMTV**P**GVVLLGSLFSR	0.764
D	FNRWFLTGMTV**D**GVVLLGSLFSR	1.647
N	FNRWFLTGMTV**N**GVVLLGSLFSR	1.062
E	FNRWFLTGMTV**E**GVVLLGSLFSR	1.178
Q	FNRWFLTGMTV**Q**GVVLLGSLFSR	1.071
K	FNRWFLTGMTV**K**GVVLLGSLFSR	1.715
R	FNRWFLTGMTV**R**GVVLLGSLFSR	1.465
H	FNRWFLTGMTV**H**GVVLLGSLFSR	0.905
F	FNRWFLTGMTV**F**GVVLLGSLFSR	-0.736
Y	FNRWFLTGMTV**Y**GVVLLGSLFSR	0.192
W	FNRWFLTGMTV**W**GVVLLGSLFSR	-0.204

Figure S6. Variations in the central position of BclxL TMD influence insertion and interaction. The insertion potential of the sequences resulting from Ala 221 (position 12 in the TMD) substitutions was calculated using the ΔG prediction server (negative ΔGs green, positive ΔGs red). b. Substitutions that permitted membrane insertion were tested (BLaTM assay, E.coli) to investigate their homo- and hetero-interaction (with the TMD of BclxL) potential. The βN T20-βC T20 homodimer was used as a negative control (gray). The βN GpA-βC GpA homodimer was used as a positive control and to normalize values across experiments. The means and standard deviations of at least three independent experiments are shown. Values above the bars indicate the level of significance (ordinary one-way ANOVA test with Dunnett

correction, p-value < 0.05) vs the T20 or BclxL TMD homo-interactions (top and bottom respectively). The level of significance (ordinary one-way ANOVA test with Dunnett correction, p-value < 0.05) when comparing the interactions between the TMD of BclxL and D1 vs D1 and D1 is included. P-values < 0.05 are highlighted in bold orange letters, p-values > 0.05 are shown in black. c. Interaction of the BclxL TMD variants with the TMD of Bcl2. The indicated chimeras (βN-βC) were co-expressed in E. coli and the resulting ampicillin LD50 was measured. The means and standard deviations of at least three independent experiments are shown. The individual value for each experiment is represented by a solid dot. The T20 homodimer was used as a negative control (gray). The GpA homodimer was used as a positive control and to normalize values across experiments. The level of significance (p-value, ordinary one-way ANOVA test with Dunnett correction) when comparing the LD50 values of the indicated interactions vs the negative control is shown above the bars; p-values < 0.05 in bold, p-values > 0.05 in black.
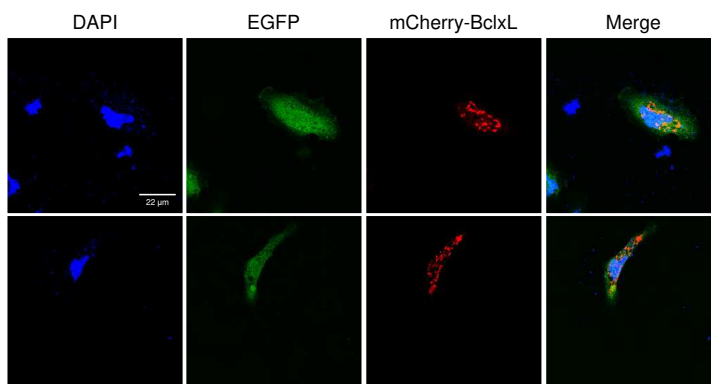
Figure S7. Subcellular localization of the eGFP. The eGFP was transfected into HeLa cells (green; n = 3) together with a mitochondrial marker (mCherry-BclxL, shown in red). DAPI staining is shown in blue. The right column of each panel shows the co-localization of both signals (yellow, visible only when the images are merged). No co-localization between eGFP and mCherry-BclxL was seen.
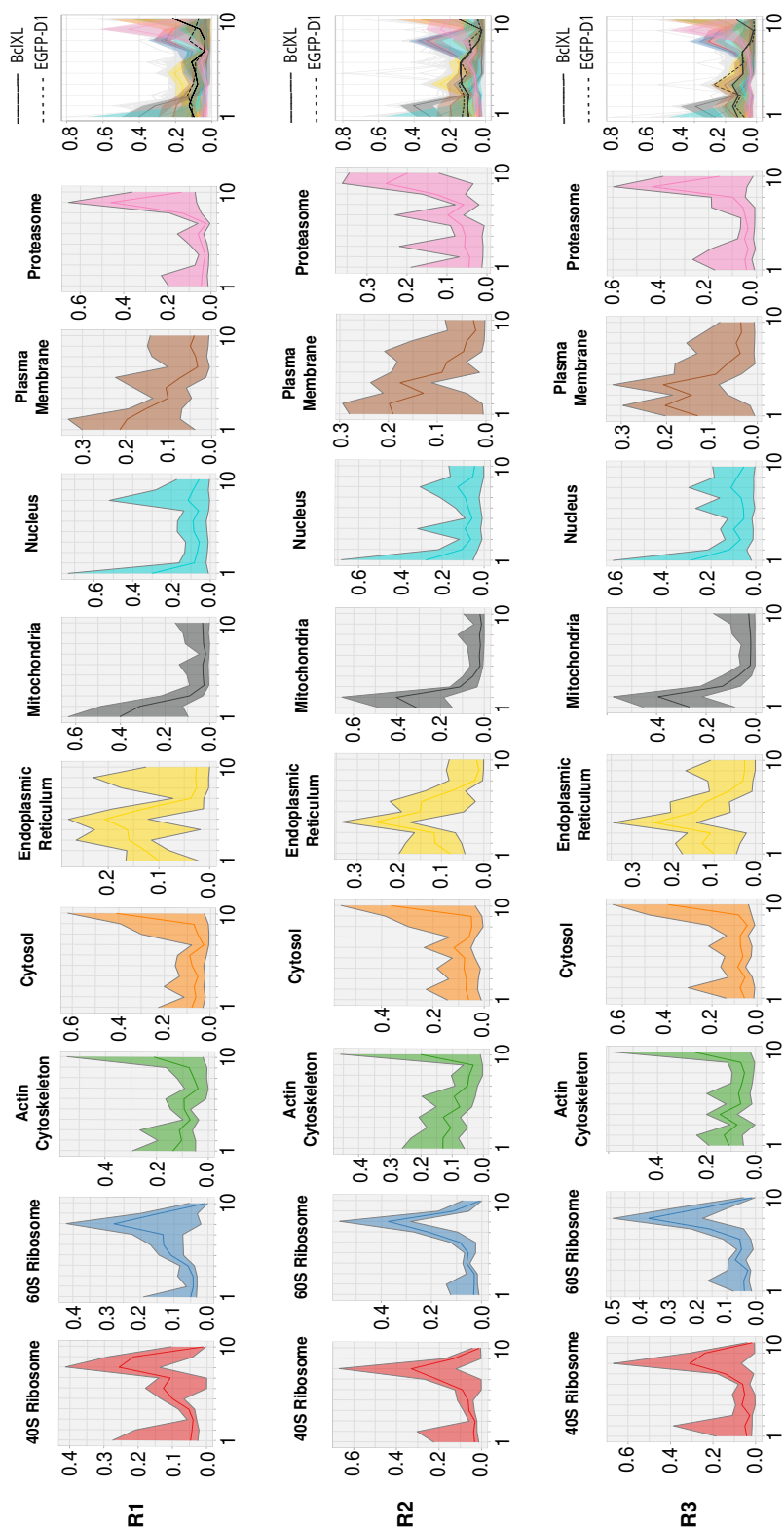
Figure S8. Abundance profiles of organelle markers. A series of differential ultracentrifugation steps was used to separate organelles and subcellular compartments. The proteins in each fraction were quantified by SWATH-MS analysis. The figure show the abundance profiles of the markers selected with the prolocGUI78 R package to identify organelles and subcellular compartments. The right panel shows the distribution of both BclxL (blue line) and eGFP-D1 (green dotted line) across fractions 1—10. The results of three independent experiments are shown (n = 3).
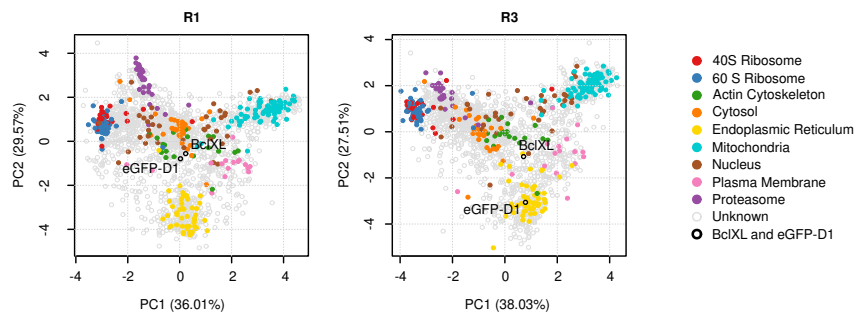
Figure S9. Localization of BclxL and eGFP-D1 by differential centrifugation. Principal Component Analysis (PCA) of the abundance profiles across fractions obtained by differential centrifugation. Each plot shows the 1962 proteins identified by SWATH-MS. The percentage in each axis represent the amount of total variability that PC1 and PC2 can explain. Organelle markers are colored. eGFP-D1 and BclxL indicated by white circles outlined in black, and the rest of the proteins ("other") are indicated by in gray circles. The plot corresponding to replicate 2 was also included in figure 6.

5. Conclusions

1. Pairs of oppositely charged residues can form salt bridges that reduce the ΔG_{app} of insertion of charged residues in TMDs.

2. SARS-CoV-2 E (envelope) protein integrates into the membrane co-translationally as a single-spanning membrane protein with Nt_{lum}/Ct_{cyt} orientation.

3. SARS-CoV-2 E protein topology is reinforced by multiple topological determinants.

4. The Ct hydrophobic region of the vBcl2s is a genuine TMD capable of interacting with cBcl2 TMDs

5. Intramembrane interactions are critical for viral and cellular Bcl2 cell fate control.

6. It is possible to design effective TMD-TMD interactions based *only* on the sequence of the target receptor.

6. Resum

L'objectiu principal d'aquesta tesi és estudiar els segments α-helicoidal de les proteïnes de membrana, des dels principis bàsics fins al disseny racional. La nostra intenció és investigar un ampli espectre de funcions desenvolupades pels dominis transmembrana (TMD) que van molt més enllà del seu paper estructural d'ancoratge de proteïnes a les membranes. L'objectiu final d'aquesta tesi és aprofundir en la nostra comprensió de la biogènesi de proteïnes de membrana des del procés d'inserció i la determinació de la topologia per caracteritzar millor aquestes proteïnes. En particular, volíem avançar en la descripció de com funcionen les interaccions TMD-TMD α-helicoidals i la seua importància en la regulació de diferents processos cel·lulars com l'apoptosi i el control de la mort cel·lular. Una millor comprensió d'aquests processos de regulació pot millorar el nostre coneixement actual i conduir a noves dianes terapèutiques. En aquesta direcció, hem explorat la inhibició de les interaccions TMD-TMD com a objectiu per regular la mort cel·lular.

Els resultats d'aquesta tesi s'han obtingut utilitzant algunes de les tècniques més rellevants utilitzades en els laboratoris de bioquímica i biologia moleculars, com per exemple: clonatge de plasmidis, mutagènesi dirigida, electroforesi SDS-PAGE, *western blotting*, expressió per transcripció/traducció *in vitro*, assajos d'apoptosi en línia cel·lular (HeLa) i també tècniques d'enginyeria de proteïnes basades en protocols de complementació bimolecular com BiFC (de l'anglès: ***Bi****molecular* ***F****luorescence* ***C****omplementation*; Complementació Bimolecular Fluorescent) i BLaTM (basat en l'ús d'una **β-La**ctamasa partida fusionada a un **TM**D).

Aquestes tècniques ens han dirigit cap a una millor descripció de com es poden inserir residus polars a les membranes biològiques. Aquestes també ens han ajudat a millorar el nostre coneixement sobre el procés d'inserció i la determinació de la topologia de proteïnes virals, així com a comprendre la complexitat de les interaccions TMD-TMD i el seu paper en el control de l'apoptosi, remarcant la importància dels TMD. Finalment, hem utilitzat aquestes interaccions per a dissenyar de manera computacional

inhibidors transmembrana com a nous possibles agents terapèutics per a regular la mort cel·lular mitjançant la modulació de les interaccions TMD-TMD. Les recerques relacionades amb aquest treball han resultat en quatre publicacions com a primer autor i un manuscrit que està sent revisat en aquest moment, els quals constitueixen els quatre capítols d'aquesta tesi, i dos articles més inclosos en la secció annexada. Els articles annexats (vegeu la *secció 8*), dels quals soc també primer autor, inclouen una revisió que descriu alguns dels mètodes més rellevants utilitzats en aquesta tesi per a l'estudi d'interaccions TMD. A més a més, la Universitat de València ha emplenat una sol·licitud de patent (P202230029) que inclou part d'aquest treball. Les pàgines que venen a continuació són un resum dels resultats més rellevants i la discussió d'aquests articles organitzats en quatre capítols.

6.1. Capítol 1: Inserció

Article:

Duart, G.*, Lamb, J.*, Ortiz-Mateu, J., Elofsson, A., Mingarro, I., 2022.
    Intra-Helical Salt Bridge Contribution to Membrane Protein Insertion. ***Journal of Molecular Biology*** 434, 167467.
    https://doi.org/10.1016/j.jmb.2022.167467
    *Equal contribution

L'objectiu d'aquest treball era estimar la contribució energètica dels ponts salins intrahèlix en la inserció de TMD en membranes biològiques. Encara no s'ha estudiat adequadament la presència de ponts de sal intrahèlix en els TMD, així com el seu impacte en la inserció. Els TMD α-helicoidals estan formats en gran part per residus apolars a causa de la naturalesa hidrofòbica de la membrana. No obstant això, en alguns casos, les proteïnes incrustades en la membrana tenen aminoàcids polars en els seus TMD amb finalitats funcionals o estructurals (Baeza-Delgado et al., 2013). De fet, la presència d'aminoàcids polars i carregats en TMD és més freqüent del que es podria esperar a causa de la naturalesa hidrofòbica de l'entorn (Bañó-Polo et al., 2012), especialment quan aquests residus es troben en parells.

Els ponts salins són interaccions electroestàtiques entre aminoàcids amb càrrega positiva i negativa que juguen un paper important en l'estabilització de proteïnes (Marqusee and Baldwin, 1987). En analitzar els TMD a partir d'estructures de proteïnes de membrana hem observat que les parelles d'aminoàcids carregats són especialment prevalents en les posicions *i, i+1*; *i, i+3* i *i, i+4*. Les parelles de càrregues amb signe contrari poden formar potencials ponts salins si estan col·locades en aquestes posicions perquè es troben totes en la mateixa cara relativa de l'hèlix i es troben suficientment a prop, en termes de distàncies atòmiques.

Per a avaluar la contribució de ponts salins a la inserció en membrana mediada pel transolocó, hem utilitzat com a vehicle la proteïna de la peptidasa líder (Lep) d'*Escherichia coli*. La proteïna Lep consta de dos TMD (H1 i H2) connectats mitjançant un llaç citosòlic (P1) i un llarg domini C-terminal (P2), que s'insereix en microsomes derivats de reticle endoplasmàtic (RE) amb els dos extrems localitzats al lumen dels microsomes (von Heijne, 1989). Els TMD dissenyats s'insereixen al domini luminal P2 flanquejats per dos llocs acceptors d'*N*-glicosilació (G1 i G2). La glicosilació ocorre exclusivament al lumen del RE (o dels microsomes) a causa de la localització del centre actiu de l'oligosacariltransferasa (un enzim associat al transolocó responsable de la transferència de grups oligosacàrids) (Braunger et al., 2018). La glicosilació d'un lloc acceptor augmenta la massa molecular aparent de la proteïna (~2.5 kDa), la qual cosa permet la seua identificació en gel d'electroforesi.

Primerament, hem començat comparant l'efecte dels residus carregats de manera oposada Lys i Asp en la inserció d'una bastida hidrofòbica artificial amb una extensió de 19 residus (L4/A15, 4 leucines i 12 alanines), dissenyada per a inserir-se de manera estable en membranes microsomals i aïllada de la seqüència flanquejant per tetrapèptids GGPG- i -GPGG (G, glicina; P, prolina) als extrems N- i C-terminal. Hem col·locat residus individuals de Lys i Asp a les posicions 8 i 12 respectivament, mentre que s'han col·locat parelles Lys-Asp cobrint les posicions 7-12 (més d'un gir complet de l'hèlix). Quan les càrregues estan presents en parelles, els nostres resultats indiquen una tendència clara a inserir-se de manera més

eficient quan aquestes es troben a distàncies que són permissives amb la formació de ponts salins (*i, i+1*; *i, i+3*; *i, i+4*), fet que no s'observa en els actuals algoritmes de predicció. Aquest resultat s'ha pogut replicar amb resultats similars utilitzant una seqüència base de Leu/Ala amb una eficiència d'inserció lleugerament major (L5/A14, 5 leucines i 4 alanines).

Després d'estudiar l'efecte dels residus amb càrrega oposada en les seqüències model d'inserció, hem decidit cercar ponts salins en proteïnes naturals. Hem analitzat els TMD a partir d'estructures d'alta resolució de proteïnes de membrana. Després, hem generat una llista de candidats potencials per a estudis posteriors en sistemes *in vitro* i *de cèl·lula completa*. Hem seleccionat l'hèlix G de l'halorhodopsina de *Natronomonas pharaonis* (codi PDB: 3QBG) i l'hèlix A de l'ATPasa de calci d'*Oryctolagus cuniculus* (codi PDB:1SU4). En ambdós casos hem estudiat la inserció d'aquestes hèlixs utilitzant assajos *in vitro* (Lep en microsomes) i en cèl·lules eucariòtiques (Lep i C_LTM).

Pel que fa a l'Halorhodopsina de *Natronomonas pharaonis*, aquesta és una proteïna que consta de 7 TMD (de l'A a la G) i un grup retinol com a cromòfor unit mitjançant una base protonada de Schiff al grup ε-amino de la Lys258, localitzada just al centre de l'hèlix G (Kanada et al., 2011). L'anàlisi *in silico* de l'estructura de l'apoproteïna 3QBG mostra una parella Lys-Arg, que involucra els residus Lys258 i Asp254 de l'hèlix G a distància relativa d'*i, i+4*. La distància entre les càrregues a l'estructura cristal·lina de l'apoproteïna és d'aproximadament 3'5 Å, una distància permissiva per a la formació d'un pont salí. A continuació, hem dissenyat tres mutants que haurien de pertorbar la formació d'un pont salí de tres maneres diferents: el mutant K258D situa dos aminoàcids carregats amb la mateixa polaritat a les posicions *i, i+4*; el mutant K258A substitueix un dels residus carregats per un aminoàcid no polar i, finalment, el doble mutant K258Y/Y259K situa les dues càrregues oposades a una distància no permissiva amb la formació d'un pot salí (*i, i+5*) mentre que manté la composició d'aminoàcids del segment. Els resultats de l'assaig de glicosilació basat en Lep indica que la seqüència salvatge així com el mutant K258A s'insereixen correctament a la membrana, però, quan el pont salí es trenca, l'eficiència d'inserció baixa de

manera substancial (0'5 kcal/mol aproximadament). Aquests resultats han sigut replicats també en cèl·lules HEK-293T. Les cèl·lules transfectades amb la quimera que conté l'hèlix G nativa han resultat pràcticament en inserció completa. No obstant això, les cèl·lules transfectades amb la construcció que presenta la seqüència *i, i+5* han resultat quasi exclusivament en translocació. Aquests resultats emfatitzen la importància dels ponts salins intrahèlix en la inserció de TMD mediada pel translocó, especialment en l'entorn cel·lular.

Pel que fa a la nostra segona proteïna d'estudi, ha sigut l'ATPasa de calci d'*Oryctolagus cuniculus*. Aquesta proteïna conté un feix de 10 hèlixs TM (de l'A a la J) (Toyoshima et al., 2000). L'anàlisi *in silico* d'1SU4 revela una parella Asp-Arg que involucra l'Asp59 i l'Arg63, al centre de l'hèlix A. La distància entre aquestes càrregues a l'estructura cristal·lina és d'aproximadament 3 Å, clarament en el rang permissiu per a la formació de ponts salins. De nou, els resultats de l'assaig de glicosilació basat en Lep demostren una inserció eficient de la seqüència nativa de l'hèlix A. No obstant això, quan la parella de residus es troba a una posició d'*i, i+5,* no permissiva amb la formació de ponts salins, l'eficiència d'inserció baixa amb un ΔG_{app} estimat d'unes 0'7 kcal/mol. En aquest cas hem observat resultats similars als experiments amb cèl·lules HEK-293T.

En resum, els residus carregats que es troben a les hèlixs α poden ser residus funcionals necessari per a desenvolupar alguna funció (Lin and Lin, 2018), però també poden exercir un paper important en l'estabilització (Armstrong and Baldwin, 1993). Les parelles de residus amb càrrega oposada poden formar ponts salins que permeten facilitar la inserció dels TMD que els contenen en exposar-se a l'entorn hidrofòbic (Whitley et al., 2021). En els TMD, les parelles de residus carregats de manera oposada són més prevalents a les posicions *i, i+1; i, i+3* i *i, i+4*, compatibles amb la formació de ponts salins. Mitjançant l'anàlisi de dos hèlixs naturals que contenen ponts salins intrahèlix hem demostrat que l'energia lliure d'inserció (ΔG_{app}) es redueix de manera significativa quan les dues càrregues oposades es troben a una distància permissiva amb la formació del pont salí. Aquests resultats indiquen que els ponts salins intrahèlix es podrien formar durant la inserció mediada pel translocó o fins i tot abans, ja que les hèlixs

transmembrana es poden compactar dins del túnel d'eixida del ribosoma (Bañó-Polo et al., 2018). La reducció del ΔG_{app} en aquestes proteïnes naturals és d'entre 0'5-0'7 kcal/mol. Com hem observat en el cas de l'hèlix G de l'halorhodopsina, aquesta reducció podria ser inclús major en el context cel·lular, ja que alguns dels components auxiliars de la maquinària d'inserció en membrana (Chitwood and Hegde, 2020; Shurtleff et al., 2018; Tamborero et al., 2011) podrien no estar presents en la preparació de vesícules microsomals. Els algoritmes actuals per a la predicció d'inserció de proteïnes de membrana tendeixen a sobreestimar la penalització d'energia lliure dels residus carregats en TMD. Incorporar l'efecte de potencials ponts salins i la conseqüent reducció del ΔG_{app} durant la inserció en membrana podria ajudar a millorar les eines de predicció futures.

6.2. Capítol 2: Topologia

Article:

Duart, G.*, García-Murria, M.J.*, Grau, B.*, Acosta-Cáceres, J.M., Martínez-Gil, L., Mingarro, I., 2020. *SARS-CoV-2 envelope protein topology in eukaryotic membranes.* **Open Biology** 10, 200209. https://doi.org/10.1098/rsob.200209
*Equal contribution

Duart, G., García-Murria, M.J., Mingarro, I., 2021a. *The SARS-CoV-2 envelope (E) protein has evolved towards membrane topology robustness.* **Biochimica et Biophysica Acta (BBA) - Biomembranes** 1863, 183608. https://doi.org/10.1016/j.bbamem.2021.183608

Tan prompte com es va fer públic el genoma del SARS-CoV-2, el virus responsable de la pandèmia global de la COVID-19, vam començar un projecte centrat en l'estudi de la proteïna de l'embolcall (E) víric. La proteïna E és la proteïna més menuda i que menys nombre de còpies presenta entre les proteïnes de membrana que es troben a l'embolcall lipídic de les partícules virals madures (Bar-On et al., 2020). Tanmateix, aquesta proteïna és crítica per a la patogènesi del SARS-CoV-2 i altres coronavirus humans (Almazán et al., 2013; Ruch and Machamer, 2012; Xia et al., 2021; Zheng et al., 2021) i ha estat descrita com una viroporina. Paga la pena

esmentar que l'sgRNA que codifica la proteïna E és un dels transcrits que s'expressa de manera més abundant, tot i la baixa quantitat de proteïna que s'expressa en els virus madurs (Wu et al., 2020). Aquest sgRNA codifica un polipèptid de 75 residus de longitud amb una massa molecular predita d'aproximadament 8 kDa. L'anàlisi comparativa de les seqüències de la proteïna E del SARS-CoV-2 i dels altres 6 coronavirus humans coneguts no mostra cap regió llarga que siga homòloga o idèntica, només la Met inicial, la Leu39, la Cys40 i la Pro54 es troben conservades de forma ubiqua. Pel que fa a la similitud general de la seqüència, la proteïna E del SARS-CoV-2 té la similitud més alta amb el SARS-CoV (94,74%) amb només diferències menors, seguida del MERS-CoV (36,00%). Curiosament, les similituds de seqüències són significativament més baixes per als altres quatre coronavirus humans, que solen causar malalties de lleus a moderades del tracte respiratori superior, típiques del refredat comú.

Per determinar la seua topologia a la membrana, hem assajat la inserció de la proteïna E en membranes microsomals utilitzant experiments de transcripció/traducció *in vitro* en presència d'aminoàcids marcats amb [^{35}S] i també en membranes eucariòtiques utilitzant cèl·lules HEK-293T. Utilitzant un assaig basat en les glicosilacions com a reporter molecular, hem determinat que la proteïna E del SARS-CoV-2 s'integra a la membrana de manera cotraduccional amb un únic TMD amb una orientació Nt_{lum}/Ct_{cyt} tant en el sistema *in vitro* com en el sistema *in vivo*. Aquesta topologia és compatible amb la capacitat de canal iònic descrita prèviament (Verdiá-Báguena et al., 2012). A més, aquesta topologia està reforçada per diferents determinants topològics codificats en la seqüència de la proteïna. En els set coronavirus humans hi ha una càrrega positiva fortament conservada situada just després de la regió hidrofòbica. Cal comentar que aquesta càrrega és una arginina (Arg38) en el MERS-CoV, el SARS-CoV i el SARS-CoV-2, mentre que en els altres coronavirus humans és una lisina. També, l'alineament de les proteïnes E del MERS-CoV, el SARS-CoV i el SARS-CoV-2 mostra una tendència a acumular un balanç de càrrega positiva a la regió C-terminal del TMD, la qual cosa es correspon amb la norma «de les càrregues positives a l'interior» (*positive-inside rule*) i suggereix un augment progressiu de la robustesa en la determinació de la topologia d'aquesta proteïna des del

MERS-CoV fins al SARS-CoV-2. Hem demostrat de manera experimental aquest augment de la robustesa mitjançant la modificació del balanç de càrregues de la proteïna sencera en les tres proteïnes E patogèniques. En totes tres, el residu d'Arg38 té un paper limitat en la determinació de la topologia. Les nostres dades també suggereixen que la mutació d'Arg a Glu present en els dos SARS-CoV a la regió N-terminal (si comparem amb la seqüència del MERS-CoV) és un dels mecanismes més probables que contribueix a l'augment comprovat de la robustesa de la determinació topològica dels SARS-CoV en convertir la càrrega neta de 0 a la regió N-terminal del MERS-CoV en una càrrega neta de -2 als dos SARS-Co. Aquest fet també concorda amb la denominada norma «de l'enriquiment de càrregues negatives a l'exterior» (*negative outside enrichment rule*) (Baker et al., 2017).

6.3. Capítol 3: Interacció i funció

Article:

García-Murria, M.J.*, **Duart, G.**.*, Grau, B., Diaz-Beneitez, E., Rodríguez, D., Mingarro, I., Martínez-Gil, L., 2020. *Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control cellular apoptosis.* **Nature Communications** 11, 6056. https://doi.org/10.1038/s41467-020-19881-9
*Equal contribution

La mort cel·lular programada és un procés fonamental per al desenvolupament dels organismes multicel·lulars que contribueix a mantindre el balanç entre mort cel·lular, proliferació i diferenciació, fet que resulta crucial per al desenvolupament de teixits i l'homeòstasi (Kerr et al., 1972). Per una altra banda, l'apoptosi també té un paper molt important en la defensa contra molts desordres, incloent-hi el càncer i malalties relacionades amb patògens, gràcies a l'eliminació selectiva de cèl·lules afectades (Häcker, 2018; Hua et al., 2019).

Un dels principals reguladors de l'apoptosi és la família de les proteïnes del limfoma de cèl·lules B, coneguda com a família Bcl2 (de

l'anglès **B-cell lymphoma 2**) (Kim et al., 2006). Les proteïnes d'aquesta família es poden dividir en antiapoptòtiques (e.g., Bcl2 i BclxL) (Boise et al., 1993), proapoptòtiques (e.g., Bax i Bak) (Oltvai et al., 1993) i BH3-only (e.g., Bid i Bmf) (Wang et al., 1996). La majoria de proteïnes proapoptòtiques i antiapoptòtiques d'aquesta família comparteixen fins a quatre dominis d'homologia amb la proteïna Bcl2, que es coneixen com a BH1, BH2, BH3 i BH4, mentre que els membres de la família BH3-only només comparteixen un domini BH3. A més, molts membres de la família Bcl2 tenen un TMD a l'extrem C-terminal que permet la inserció efectiva de la proteïna en la bicapa lipídica adequada (Delbridge et al., 2016).

Les proteïnes Bcl2 cel·lulars (cBcl2) poden interaccionar físicament les unes amb les altres per formar homooligòmers i heterooligòmers que resulten fonamentals per a la regulació de la mort cel·lular programada (Cosentino and García-Sáez, 2017; Kelekar et al., 1997; Oltvai et al., 1993; O'Neill et al., 2006; Wang et al., 1996; Xie et al., 1998). Per tal de previndre la mort prematura de les cèl·lules de l'hoste, molts virus han desenvolupat homòlegs funcionals de les cBcl2, coneguts com Bcl2 virals (vBcl2), com a estratègia per modular la mort cel·lular (Kvansakul et al., 2017; Polčic et al., 2017). Tot i que existeix poca homologia de seqüències entre les vBcl2 i les cBcl2, les estructures cristal·lines mostren homologia estructural en dominis clau (Galluzzi et al., 2008; Kvansakul and Hinds, 2013).

En aquest treball, hem començat comprovant que les vBcl2 contenen un TMD funcional al seu extrem C-terminal, així com ocorre amb les cBcl2. Hem seleccionat sis proteïnes vBcl2 de dues famílies virals diferents (3 herpesvirus i 3 poxvirus). BHRF1 (*gammaherpesvirus humà 4 – virus d'Epstein Bar*, HHV4) (Pearson et al., 1987), ORF16 (*gammaherpesvirus humà 8* – herpesvirus associat al sarcoma de Kaposi, HHV8) (Cheng et al., 1997), ORF16 (*gammaherpesvirus Boví 4*, BoHV4) (Bellows et al., 2000), F1L (*Vaccinia virus*, VacV) (Nichols et al., 2017; Wasilenko et al., 2003), M11L (*Myxoma virus*, MyxV) (Douglas et al., 2007; Nichols et al., 2017) i ORFV125 (*Orf virus*, OrfV) (Westphal et al., 2007). Amb la finalitat d'evitar confusió, a partir d'ací utilitzarem l'acrònim viral per a referir-nos a les proteïnes vBcl2 esmentades. L'anàlisi *in silico*

suggereix la presència de TMD en les vBcl2 seleccionades. A continuació, l'objectiu ha sigut explorar la capacitat d'inserció en membrana dels segments predits utilitzant un assaig *in vitro* que fa ús de la peptidasa líder (Lep) d'*E. coli* descrit anteriorment (vegeu la *secció 6.1*). Utilitzant aquest sistema basat en glicosilacions hem determinat que totes les regions estudiades de les vBcl2 s'insereixen eficientment en membranes microsomals.

Després de determinar que les vBcl2 inclouen un TMD d'ancoratge a la regió C-terminal, hem volgut estudiar quin paper tenen aquestes regions més enllà de la funció d'àncora estructural. Com que s'han reportat interaccions entre els TMD de les cBcl2 en membranes biològiques (Andreu-Fernández et al., 2017), hem decidit estudiar la capacitat d'homooligomerització i heterooligomerització dels TMD de les vBcl2 en membranes biològiques. Amb aquesta finalitat, hem utilitzant dos sistemes de complementació bimolecular basats en reporters i organismes model diferents. Per començar, hem utilitzat el sistema de BiFC (de l'anglès **B**imolecular **F**luorescent **C**omplementation) (Kerppola, 2006) adaptat per a l'estudi d'interaccions intramembrana en cèl·lules eucariòtiques. Aquesta tècnica es basa en l'ús d'una proteïna venus fluorescent (VFP) partida. Cadascun dels dos fragments no-fluorescents de la VFP es fusiona a un dels TMD que s'estudien i s'expressen en cèl·lules eucariotes. Els dos fragments de la VFP no presenten afinitat intrínseca l'un per l'altre. La VFP es reconstituirà (i la seua fluorescència) només quan es reporte una interacció TMD-TMD. Els TMD d'HHV4, HHV8, VacV, MyxV i OrfV han mostrat capacitat d'homointeraccionar per sobre dels controls i de manera similar a l'observada en el TMD de cBcl2. Tot i això, BoHV no ha mostrat fluorescència associada a VFP significativament major als controls negatius. L'anàlisi per *western blot* mostra nivells d'expressió comparables per a totes les quimeres.

Per tal d'investigar les possibles interaccions TMD-TMD heteromèriques entre vBcl2 i cBcl3 hem utilitzat el sistema de BiFC, descrit anteriorment. Hem estudiat les interaccions potencials entre les vBcl2 i les proteïnes antiapoptòtiques cel·lulars Bcl2 i BclxL. Val a dir que tots els

TMD inclosos en l'assaig han mostrat capacitat d'interacció amb Bcl2. No obstant això, tot i que la majoria dels TMD de les vBcl2 també han mostrat interacció amb el TMD de BclxL, BoHV i MyxV no n'hi han mostrat. També hem analitzat les interaccions entre els TMD de les vBcl2 i els TMD de les proteïnes cel·lulars proapoptòtiques Bax i Bak. En aquest cas, tots tres TMD poxvirals (VacV, MyxV i OrfV) han mostrat capacitat d'interacció amb Bax i Bak en absència d'estímul apoptòtic i de qualsevol contacte establert a partir dels dominis solubles. En contraposició, els TMD dels herpesvirus HHV4, HHV8 i BoHV no han mostrat cap interacció amb els TMD de Bax i Bak. Finalment, també hem afegit a l'*screening* de BiFC els TMD dels moduladors de l'apoptosi BH3-only Bik i Bmf (Andreu-Fernández et al., 2016). Els TMD d'HHV4, HHV8, VacV i OrfV poden interaccionar amb el TMD de Bik. Tanmateix, les interaccions virals amb el TMD de Bmf són més limitades, amb HHV8 com a únic TMD capaç d'establir interacció intramembrana amb els TMD de Bmf. També, hem decidit explorar amb més profunditat algunes d'aquestes interaccions (e.g., HHV8-Bcl2 i MyxV-Bax) utilitzant l'assaig de BLaTM (Schanzenbach et al., 2017) i modelització computacional. BLaTM és una eina genètica dissenyada per a l'estudi d'interaccions TMD-TMD en membranes bacterianes. Aquesta eina està basada en l'ús d'una β-Lactamasa partida fusionada als TMD d'estudi i expressada en cèl·lules d'*E. coli*. Les interaccions TMD-TMD dirigeixen la reconstitució de l'activitat β-Lactamasa, la qual confereix resistència a l'ampicil·lina. En aquest assaig, l'LD_{50} de l'antibiòtic serveix com un indicador de la força de la interacció. Cal remarcar que els bacteris, en aquest cas, únicament poden créixer en presència d'ampicil·lina quan la β-Lactamasa es troba reconstituïda al periplasma. Com a conseqüència, l'assaig de BLaTM també reporta la inserció en la membrana de les regions assajades. Amb aquest sistema, hem tornat a assajar d'una manera més quantitativa algunes de les interaccions estudiades anteriorment. En línies generals, els nostres descobriments assenyalen una intricada xarxa d'interaccions entre els TMD d'origen viral i cel·lular.

Finalment, hem volgut determinar si aquesta xarxa d'interaccions TMD-TMD descrita juga un paper important en la modulació de l'apoptosi.

Amb aquesta finalitat, hem transfectat cèl·lules HeLa amb Bcl2, HHV8 o MyxV amb o sense el TMD (variants FL i ΔTM, respectivament). També hem inclòs quimeres en les quals hem substituït el TMD de cadascuna de les proteïnes descrites pel TMD de la proteïna mitocondrial no-apoptòtica TOMM20 (Bcl2-T20, HHV8-T20 i MyxV-T20, respectivament). En assajos previs hem pogut observar que el TMD d'aquesta proteïna (T20) no és capaç d'interaccionar amb el TMD de cap proteïna Bcl2, ni viral ni cel·lular. Les cèl·lules han sigut transfectades amb les construccions pertinents i posteriorment tractades amb doxorubicina, una droga inductora de l'apoptosi (Rooswinkel et al., 2014). Com era d'esperar, les proteïnes FL han sigut capaces de previndre l'apoptosi. En canvi, quan el TMD ha sigut eliminat, cap de les proteïnes ha mantingut la seua capacitat antiapoptòtica. De manera similar, les quimeres que contenen el TMD de T20 tampoc han pogut controlar l'apoptosi induïda per doxorubicina, tot i que es localitzen en la mateixa membrana biològica que les proteïnes FL. Aquests experiments també han sigut replicats utilitzant estímuls apoptòtics diferents: apoptosi induïda per virus i apoptosi induïda per Bax (només les variants FL i T20 de Bcl2 i MyxV).

En resum, hem identificat la regió hidrofòbica C-terminal de les proteïnes vBcl2 com autèntics TMD amb capacitat d'interaccionar amb els TMD de les cBcl2. També hem demostrat que aquestes interaccions intramembrana són crucials per al control viral del destí cel·lular. Aquest estudi proporciona una comprensió més profunda de com els virus controlen la mort cel·lular per al seu propi benefici i també contribueix a explicar com els virus interaccionen amb els seus hostes.

6.4. Capítol 4: Disseny d'interaccions entre proteïnes de membrana

Article:

Duart, G., Elazar, A., Weinsten, J, Gadea-Salom, L., Ortiz-Mateu, J., Fleishman, S., Mingarro, I., Martínez-Gil, L. *Computational design of BclxL inhibitors that target transmembrane domain interactions. **Under revision.***

En aquest capítol, hem tractat d'explorar les interaccions proteïna-proteïna (PPI) intramembrana de la proteïna BclxL, com també el seu rol en la funció antiapoptòtica. Hem utilitzat la informació obtinguda per a dissenyar de manera computacional inhibidors capaços de segrestar de manera selectiva el TMD de Bclxl per tal de revertir la resistència a l'apoptosi.

Tal com hem explicat en la *secció 6.3*, els membres de la família de proteïnes Bcl2 poden interaccionar els uns amb els altres, formant homooligòmers i heterooligòmers (Andreu-Fernández et al., 2017; Cosentino and García-Sáez, 2017; Kelekar et al., 1997; Oltvai et al., 1993; Wang et al., 1996). Aquestes PPI són part d'una important xarxa de regulació que controla la mort cel·lular programada mitjançant la permeabilització de la membrana externa mitocondrial. En cèl·lules sanes, els membres antiapoptòtics de la família Bcl2 inhibeixen l'activació de les proteïnes proapoptòtiques mitjançant interacció directa o pel segrest de proteïnes BH3-only (Kim et al., 2006). Sota un estímul apoptòtic, les proteïnes proapoptòtiques i BH3-only s'alliberen i poden induir la permeabilització de la membrana externa mitocondrial. Fins ara, es creia que les interaccions entre els membres de la família Bcl2 ocorrien de manera exclusiva a través de les regions solubles, especialment a través dels dominis BH (Dadsena et al., 2021). No obstant això, troballes recents han demostrat que els seus TMD també participen d'aquestes PPI (Andreu-Fernández et al., 2017; Lucendo et al., 2020) i, tal com s'ha comprovat en el Capítol 3, aquestes interaccions intramembrana resulten fonamentals per al control de l'apoptosi.

Entre les proteïnes antiapoptòtiques, *Bcl-2 like protein 1*, més coneguda com a BclxL, exhibeix funcions importants en nombrosos tipus de càncer. En melanoma, BclxL participa en l'acció d'evitar que les cèl·lules executen l'apoptosi, indueix resistència a fàrmacs i promou la migració cel·lular, la invasió i l'angiogènesi (Lucianò et al., 2021). A causa de la rellevància de BclxL en la progressió del càncer, s'han estudiat diferents estratègies per tal d'inhibir-la (Lucianò et al., 2021) però es necessiten encara inhibidors nous menys adversos.

Primer, hem volgut determinar si el TMD de BclxL pot interaccionar amb altres membres proapoptòtics i antiapoptòtics de la família Bcl2. Hem estudiat els potencials contactes intramembrana utilitzant el sistema de BLaTM (veure *secció 6.3*). Utilitzant aquest assaig, hem posat a prova l'homooligomerització del TMD de BclxL, així com les possibles interaccions heterooligomèriques amb el TMD dels membres antiapoptòtics (Bcl2), proapoptòtics (Bax i Bak) i BH3-only (Bik i Bmf) de la família Bcl2. Pel que fa als nostres resultats, hem pogut determinar que el TMD de BclxL forma homooligòmers dèbils. A més, hem identificat heterooligòmers transmembrana amb Bcl2, Bax i Bak. Per tal de verificar aquesta capacitat d'interaccionar del TMD de BclxL en cèl·lules eucariòtiques, hem utilitzat l'assaig de BiFC (Kerppola, 2006), adaptat per a l'estudi d'interaccions intramembrana (Andreu-Fernández et al., 2017; Grau et al., 2017) (vegeu *secció 6.3*). Els nostres resultats indiquen que el TMD de BclxL pot homooligomeritzar i heterooligomeritzar amb els TMD de Bcl2, Bak i Bik en membranes eucariotes.

Després, per tal d'investigar si aquestes interaccions TMD-TMD són necessàries per a la regulació cel·lular de l'apoptosi efectuada per BclxL, hem transfectat cèl·lules HeLa amb BclxL sencera amb o sense la regió transmembrana (BclxL-FL i BclxL-ΔTMD, respectivament). De manera addicional hem inclòs una quimera en la qual el TMD de BclxL s'ha substituït pel TMD de TOMM20 (BclxL-T20). Els nostres resultats indiquen que, quan s'utilitza doxorubicina com a estímul per a induir la mort cel·lular, BclxL necessita el seu TMD per bloquejar l'apoptosi.

Com que BclxL es localitza principalment al mitocondri (Fang et al., 1994; González-García et al., 1994; Zamzami et al., 1998) i a causa de la importància del TMD per a la localització de les proteïnes de membrana (Martínez-Gil et al., 2011), hem explorat com afecta a la localització subcel·lular la deleció o la substitució de la regió hidrofòbica C-terminal de BclxL. Per analitzar la localització, hem expressat BclxL (BclxL-FL), BclxL-T20 i BclxL-ΔTMD en cèl·lules HeLa en paral·lel amb un marcador mitocondrial fluorescent. La imatges de microscòpia de fluorescència

confocal revelen que tant BclxL-FL com BclxL-T20 es localitzen al mitocondri, mentre que la variant BclxL-ΔTMD, com es podria esperar, mostra una distribució citosòlica.

Una vegada demostrat que les interaccions TMD-TMD són crucials per al control de la funció antiapoptòtica de BclxL, hem tractat de dissenyar un inhibidor específic per a aquestes PPI intramembrana. El disseny de l'inhibidor ha començat amb la modelització de l'homooligòmer TMD de BclxL utilitzant el programari TMHOP (*Trans-membrane Homo Oligomer Predictor*; Predictor d'homooligòmers transmembrna) (Weinstein et al., 2019). TMHOP utilitza simulacions de Rosetta simètriques *ab initio* de plegament-i-acoblament per a tots els àtoms en un entorn implícit de membrana que prediu milers de conformacions de baixa energia basant-se en una funció energètica que recau en mesures empíriques de propensió d'inserció dels diferents aminoàcids (Elazar et al., 2016; Weinstein et al., 2019). Basant-nos en les característiques estructurals i l'energia de Rosetta associada, hem seleccionat un model de TMHOP que forma un dímer paral·lel ben empaquetat. El model seleccionat s'ha introduït en l'algoritme de disseny FuncLib (Khersonsky et al., 2018) per a generar seqüències de més alta afinitat que puguen utilitzar-se com inhibidors. L'algoritme de FuncLib utilitza càlculs de disseny de Rosetta per a enumerar les combinacions de substitucions d'aminoàcids tolerades en posicions específiques. Després, relaxa cada combinació utilitzant minimització de proteïna completa (basada en la funció d'energia de Rosetta per a membrana) (Weinstein et al., 2019) i ordena aquestes combinacions per energia. Com que és ben sabut que els TMD individuals són propensos a autoassociar-se (Grau et al., 2017), hem dissenyat les seqüències minimitzant el potencial d'homooligomerització mitjançant les següents normes: i) volem dissenyar una seqüència capaç d'interaccionar amb el TMD natural de BclxL. Amb aquesta finalitat, hem utilitzat selecció positiva per a heterodímers (FuncLib no-simètric; $\Delta\Delta G$<+1 unitats d'energia de Rosetta; R.e.u). ii) Hem utilitzat selecció negativa per a un nou homodímer (FuncLib simètric; $\Delta\Delta G$>+5 R.e.u). Com a conseqüència, l'algoritme inclourà només diversificació de seqüència que permeta a la nova seqüència interaccionar de forma eficient amb el TMD natiu de BclxL però que evite que la nova

seqüència puga interaccionar amb ella mateixa. Aquest procés ha resultat en tres seqüències dissenyades de potencials inhibidors TMD (anomenats D1, D2 i D3).

Hem verificat que aquestes seqüències dissenyades poden inserir-se correctament en microsomes derivats de RE utilitzant un assaig de transcripció/traducció *in vitro*. Després, utilitzant BLaTM, hem analitzat les interaccions entre el TMD de Bclxl i els inhibidors dissenyats: D1, D2 i D3. Els resultats d'aquests experiments mostren que D1 pot unir-se de manera eficient al TMD de BclxL sense formar homooligòmers, tal com ho havíem planificat en el procés de disseny. Cal destacar que la interacció entre el TMD de BclxL i D1 és més forta que la de l'homooligòmer natural del TMD de BclxL. Per una altra banda, malgrat que D2 i D3 no formen homooligòmers, tampoc interaccionen significativament amb el TMD de BclxL. També, hem estudiat l'especificitat de la interacció observada entre D1 i BclxL. Per determinar aquesta especificitat d'interacció, hem posat a prova l'associació entre D1 i el TMD de Bcl2, una altra proteïna antiapoptòtica. No hem detectat cap interacció entre D1 i el TMD de Bcl2. Com a conseqüència, qualsevol efecte de D1 sobre la supervivència cel·lular derivarà molt probablement de la seua interacció amb el TMD de BclxL. Addicionalment, hem utilitzat el sistema de BiFC per a assegurar que les interaccions entre D1 i el TMD de BclxL es mantenen en membranes eucariòtiques. Els resultats indiquen que D1 pot unir el TMD de BclxL de manera eficient mentre que no forma homooligòmers en cèl·lules eucariotes.

Per tal d'inhibir l'efecte antiapoptòtic de BclxL, les seqüències dissenyades han de localitzar-se en el mateix compartiment cel·lular on es troba BclxL. A fi d'estudiar la seua localització, hem fusionat les seqüències de D1 i D2 a l'extrem C-terminal de la proteïna fluorescent eGFP (eGFP-D1 i eGFP-D2) i hem expressat aquestes construccions en cèl·lules HeLa conjuntament amb BclxL fusionada a la proteïna fluorescent mCherry (mCherry-BclxL). Posteriorment, hem analitzat la localització subcel·lular d'aquestes quimeres per microscòpia confocal de fluorescència. Tant eGFP-D1 com eGFP-D2 han mostrat una col·localització robusta amb el marcador mCherry-BclxL. Per assegurar que D1 i BclxL coexisteixen en el mateix

compartiment cel·lular, hem realitzat un segon assaig de localització que es basa en l'ús d'ultracentrifugacions diferencials d'orgànuls (Geladaki et al., 2019). Després, hem analitzat el contingut proteic de cadascuna de les fraccions mitjançant *sequential window acquisition of all theoretical mass spectra* (SWATH-MS) (Rotello and Veenstra, 2021; Zhang et al., 2020). BclxL i eGFP-D1 tenen una distribució de perfils similar que suggereix una localització subcel·lular també similar.

Finalment, hem estudiat l'efecte antiapoptòtic de D1 i D2. Hem transfectat cèl·lules HeLa amb BclxL juntament amb les quimeres eGFP-T20, eGFP-D1, eGFP-D2 i eGFP-xL. Com a control hem utilitzat també cèl·lules que no s'han transfectat amb BclxL ni cap de les quimeres però s'han transfectat amb plasmidi buit per mantenir constant la quantitat de DNA transfectat a les diferents mostres. Després de la transfecció, hem tractat les cèl·lules amb doxorubicina per induir l'apoptosi. Les cèl·lules transfectades amb eGFP-T20 o eGFP-xL més BclxL han mostrat un bloqueig eficient de l'apoptosi induïda per la doxorubicina. Remarcablement, la transfecció d'eGFP-D1 ha eliminat totalment l'efecte antiapoptòtic de BclxL. D2 també és capaç de reduir la viabilitat cel·lular però de manera menys dràstica que D1. Cal remarcar que no s'observen diferències significatives entre les mostres transfectades amb eGFP-D1 o eGFP-D2 (més BclxL), i les cèl·lules transfectades amb plasmidi buit tractades en tots els casos amb doxorubicina. Aquest resultat indica que D1 i D2 poden inhibir l'efecte antiapoptòtic de BclxL fins al nivell de no transfectar la proteïna. També hem estudiat la viabilitat cel·lular després de la transfecció i hem determinat que l'expressió de D1 i D2 no resulta tòxica per a les cèl·lules HeLa, una característica fonamental per al disseny d'inhibidors no-tòxics.

En resum, aquests resultats aporten evidències sobre la importància de les interaccions TMD-TMD sobre el control de l'apoptosi, particularment en el cas de BclxL. També, hem dissenyat de manera satisfactòria seqüències amb la capacitat d'inhibir específicament l'acció antiapoptòtica de BclxL. Aquest treball mostra una via per al disseny d'inhibidors efectius basats únicament en la seqüència de la diana receptora. El fet que dues de les tres seqüències dissenyades hagen mostrat l'efecte i les interaccions desitjades

subratlla la precisió de l'estratègia de modelatge de TMHOP i el disseny de l'algoritme de FuncLib, els quals han sigut utilitzats per a una gran varietat de tasques de disseny en proteïnes solubles. Aquest treball millora de manera significativa la nostra comprensió sobre el reconeixement específic de seqüències en membranes i obre un camí per a una nova generació de medicaments contra el càncer.

7. Bibliography

Allen, K.N., Entova, S., Ray, L.C., Imperiali, B., 2019. Monotopic Membrane Proteins Join the Fold. Trends Biochem Sci 44, 7–20. https://doi.org/10.1016/j.tibs.2018.09.013

Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37, 420–423. https://doi.org/10.1038/s41587-019-0036-z

Almazán, F., DeDiego, M.L., Sola, I., Zuñiga, S., Nieto-Torres, J.L., Marquez-Jurado, S., Andrés, G., Enjuanes, L., 2013. Engineering a Replication-Competent, Propagation-Defective Middle East Respiratory Syndrome Coronavirus as a Vaccine Candidate. mBio 4. https://doi.org/10.1128/mBio.00650-13

Almén, M.S., Nordström, K.J., Fredriksson, R., Schiöth, H.B., 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, 50. https://doi.org/10.1186/1741-7007-7-50

Andreu-Fernández, V., García-Murria, M.J., Bañó-Polo, M., Martin, J., Monticelli, L., Orzáez, M., Mingarro, I., 2016. The C-terminal Domains of Apoptotic BH3-only Proteins Mediate Their Insertion into Distinct Biological Membranes *. Journal of Biological Chemistry 291, 25207–25216. https://doi.org/10.1074/jbc.M116.733634

Andreu-Fernández, V., Sancho, M., Genovés, A., Lucendo, E., Todt, F., Lauterwasser, J., Funk, K., Jahreis, G., Pérez-Payá, E., Mingarro, I., Edlich, F., Orzáez, M., 2017. Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. Proc Natl Acad Sci USA 114, 310–315. https://doi.org/10.1073/pnas.1612322114

Armstrong, K.M., Baldwin, R.L., 1993. Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges. Proceedings of the National Academy of Sciences 90, 11337–11340. https://doi.org/10.1073/pnas.90.23.11337

Baeza-Delgado, C., Marti-Renom, M.A., Mingarro, I., 2013. Structure-based statistical analysis of transmembrane helices. Eur Biophys J 42, 199–207. https://doi.org/10.1007/s00249-012-0813-9

Baeza-Delgado, C., von Heijne, G., Marti-Renom, M.A., Mingarro, I., 2016. Biological insertion of computationally designed short transmembrane segments. Sci. Rep. 6, 23397.

Baker, J.A., Wong, W.-C., Eisenhaber, B., Warwicker, J., Eisenhaber, F., 2017. Charged residues next to transmembrane regions revisited: "Positive-inside rule" is complemented by the "negative inside depletion/outside enrichment rule." BMC Biology 15, 66. https://doi.org/10.1186/s12915-017-0404-4

Bañó-Polo, M., Baeza-Delgado, C., Orzáez, M., Marti-Renom, M.A., Abad, C., Mingarro, I., 2012. Polar/Ionizable Residues in Transmembrane

Segments: Effects on Helix-Helix Packing. PLOS ONE 7, e44263. https://doi.org/10.1371/journal.pone.0044263

Bañó-Polo, M., Baeza-Delgado, C., Tamborero, S., Hazel, A., Grau, B., Nilsson, I., Whitley, P., Gumbart, J.C., von Heijne, G., Mingarro, I., 2018. Transmembrane but not soluble helices fold inside the ribosome tunnel. Nat Commun 9, 5246. https://doi.org/10.1038/s41467-018-07554-7

Bañó-Polo, M., Martínez-Gil, L., Wallner, B., Nieva, J.L., Elofsson, A., Mingarro, I., 2013. Charge pair interactions in transmembrane helices and turn propensity of the connecting sequence promote helical hairpin insertion. J Mol Biol 425, 830–840. https://doi.org/10.1016/j.jmb.2012.12.001

Bar-On, Y.M., Flamholz, A., Phillips, R., Milo, R., 2020. SARS-CoV-2 (COVID-19) by the numbers. eLife 9, e57309. https://doi.org/10.7554/eLife.57309

Baum, D.A., Baum, B., 2014. An inside-out origin for the eukaryotic cell. BMC Biol 12, 76. https://doi.org/10.1186/s12915-014-0076-2

Bellows, D.S., Chau, B.N., Lee, P., Lazebnik, Y., Burns, W.H., Hardwick, J.M., 2000. Antiapoptotic Herpesvirus Bcl-2 Homologs Escape Caspase-Mediated Conversion to Proapoptotic Proteins. J Virol 74, 5024–5031.

Bogdanov, M., Dowhan, W., Vitrac, H., 2014. Lipids and topological rules governing membrane protein assembly. Biochim Biophys Acta 1843, 1475–1488. https://doi.org/10.1016/j.bbamcr.2013.12.007

Boise, L.H., González-García, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nuñez, G., Thompson, C.B., 1993. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell 74, 597–608. https://doi.org/10.1016/0092-8674(93)90508-n

Braunger, K., Pfeffer, S., Shrimal, S., Gilmore, R., Berninghausen, O., Mandon, E.C., Becker, T., Förster, F., Beckmann, R., 2018. Structural basis for coupling protein transport and N-glycosylation at the mammalian endoplasmic reticulum. Science 360, 215–219. https://doi.org/10.1126/science.aar7899

Brosig, B., Langosch, D., 1998. The dimerization motif of the glycophorin A transmembrane segment in membranes: importance of glycine residues. Protein Sci 7, 1052–1056. https://doi.org/10.1002/pro.5560070423

Casares, D., Escribá, P.V., Rosselló, C.A., 2019. Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues. Int J Mol Sci 20, 2167. https://doi.org/10.3390/ijms20092167

Cheng, E.H., Nicholas, J., Bellows, D.S., Hayward, G.S., Guo, H.G., Reitz, M.S., Hardwick, J.M., 1997. A Bcl-2 homolog encoded by Kaposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. Proc Natl Acad Sci U S A 94, 690–694. https://doi.org/10.1073/pnas.94.2.690

Chitwood, P.J., Hegde, R.S., 2020. An intramembrane chaperone complex facilitates membrane protein biogenesis. Nature 584, 630–634. https://doi.org/10.1038/s41586-020-2624-y

Choma, C., Gratkowski, H., Lear, J.D., DeGrado, W.F., 2000. Asparagine-mediated self-association of a model transmembrane helix. Nat Struct Biol 7, 161–166. https://doi.org/10.1038/72440

Corin, K., Bowie, J.U., 2022. How physical forces drive the process of helical membrane protein folding. EMBO Rep 23, e53025. https://doi.org/10.15252/embr.202153025

Cosentino, K., García-Sáez, A.J., 2017. Bax and Bak Pores: Are We Closing the Circle? Trends Cell Biol 27, 266–275. https://doi.org/10.1016/j.tcb.2016.11.004

Cymer, F., von Heijne, G., White, S.H., 2015. Mechanisms of integral membrane protein insertion and folding. J Mol Biol 427, 999–1022. https://doi.org/10.1016/j.jmb.2014.09.014

Dadsena, S., King, L.E., García-Sáez, A.J., 2021. Apoptosis regulation at the mitochondria membrane level. Biochim Biophys Acta Biomembr 1863, 183716. https://doi.org/10.1016/j.bbamem.2021.183716

Dawson, J.P., Weinger, J.S., Engelman, D.M., 2002. Motifs of serine and threonine can drive association of transmembrane helices. J Mol Biol 316, 799–805. https://doi.org/10.1006/jmbi.2001.5353

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 Sci 12, 647–665. https://doi.org/10.1110/ps.0236503

Delbridge, A.R.D., Grabow, S., Strasser, A., Vaux, D.L., 2016. Thirty years of BCL-2: Translating cell death discoveries into novel cancer therapies. Nature Reviews Cancer 16, 99–109. https://doi.org/10.1038/nrc.2015.17

Do, H., Falcone, D., Lin, J., Andrews, D.W., Johnson, A.E., 1996. The cotranslational integration of membrane proteins into the phospholipid bilayer is a multistep process. Cell 85, 369–378. https://doi.org/10.1016/s0092-8674(00)81115-0

Douglas, A.E., Corbett, K.D., Berger, J.M., McFadden, G., Handel, T.M., 2007. Structure of M11L: A myxoma virus structural homolog of the apoptosis inhibitor, Bcl-2. Protein Sci 16, 695–703. https://doi.org/10.1110/ps.062720107

Edidin, M., 2003. Lipids on the frontier: a century of cell-membrane bilayers. Nat Rev Mol Cell Biol 4, 414–418. https://doi.org/10.1038/nrm1102

Elazar, A., Chandler, N.J., Davey, A.S., Weinstein, J.Y., Nguyen, J.V., Trenker, R., Cross, R.S., Jenkins, M.R., Call, M.J., Call, M.E., Fleishman, S.J., 2022. De novo-designed transmembrane domains tune engineered receptor functions. Elife 11, e75660. https://doi.org/10.7554/eLife.75660

Elazar, A., Weinstein, J., Biran, I., Fridman, Y., Bibi, E., Fleishman, S.J., 2016. Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. eLife 5, e12125. https://doi.org/10.7554/eLife.12125

Fang, W., Rivard, J.J., Mueller, D.L., Behrens, T.W., 1994. Cloning and molecular characterization of mouse bcl-x in B and T lymphocytes. The Journal of Immunology 153, 4388–4398.

Finger, C., Escher, C., Schneider, D., 2009. The single transmembrane domains of human receptor tyrosine kinases encode self-interactions. Sci Signal 2, ra56. https://doi.org/10.1126/scisignal.2000547

Galluzzi, L., Brenner, C., Morselli, E., Touat, Z., Kroemer, G., 2008. Viral control of mitochondrial apoptosis. PLoS Pathog 4, e1000018. https://doi.org/10.1371/journal.ppat.1000018

Gavel, Y., von Heijne, G., 1992. The distribution of charged amino acids in mitochondrial inner-membrane proteins suggests different modes of membrane integration for nuclearly and mitochondrially encoded proteins. Eur J Biochem 205, 1207–1215. https://doi.org/10.1111/j.1432-1033.1992.tb16892.x

Geladaki, A., Kočevar Britovšek, N., Breckels, L.M., Smith, T.S., Vennard, O.L., Mulvey, C.M., Crook, O.M., Gatto, L., Lilley, K.S., 2019. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. Nat Commun 10, 331. https://doi.org/10.1038/s41467-018-08191-w

Goder, V., Crottet, P., Spiess, M., 2000. In vivo kinetics of protein targeting to the endoplasmic reticulum determined by site-specific phosphorylation. EMBO J 19, 6704–6712. https://doi.org/10.1093/emboj/19.24.6704

Goder, V., Spiess, M., 2001. Topogenesis of membrane proteins: determinants and dynamics. FEBS Lett 504, 87–93. https://doi.org/10.1016/s0014-5793(01)02712-0

González-García, M., Pérez-Ballestero, R., Ding, L., Duan, L., Boise, L.H., Thompson, C.B., Núñez, G., 1994. bcl-XL is the major bcl-x mRNA form expressed during murine development and its product localizes to mitochondria. Development 120, 3033–3042.

Görlich, D., Hartmann, E., Prehn, S., Rapoport, T.A., 1992. A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature 357, 47–52. https://doi.org/10.1038/357047a0

Gorter, E., Grendel, F., 1925. ON BIMOLECULAR LAYERS OF LIPOIDS ON THE CHROMOCYTES OF THE BLOOD. J Exp Med 41, 439–443.

Grau, B., Javanainen, M., García-Murria, M.J., Kulig, W., Vattulainen, I., Mingarro, I., Martínez-Gil, L., 2017. The role of hydrophobic matching on transmembrane helix packing in cells. Cell Stress 1, 90–106. https://doi.org/10.15698/cst2017.11.111

Guna, A., Hegde, R.S., 2018. Transmembrane Domain Recognition during Membrane Protein Biogenesis and Quality Control. Curr Biol 28, R498–R511. https://doi.org/10.1016/j.cub.2018.02.004

Guna, A., Volkmar, N., Christianson, J.C., Hegde, R.S., 2018. The ER membrane protein complex is a transmembrane domain insertase. Science 359, 470–473. https://doi.org/10.1126/science.aao3099

Gupta, A., Becker, T., 2021. Mechanisms and pathways of mitochondrial outer membrane protein biogenesis. Biochim Biophys Acta Bioenerg 1862, 148323. https://doi.org/10.1016/j.bbabio.2020.148323

Gurezka, R., Laage, R., Brosig, B., Langosch, D., 1999. A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial

transmembrane segments. J Biol Chem 274, 9265–9270. https://doi.org/10.1074/jbc.274.14.9265

Häcker, G., 2018. Apoptosis in infection. Microbes Infect 20, 552–559. https://doi.org/10.1016/j.micinf.2017.10.006

Harayama, T., Riezman, H., 2018. Understanding the diversity of membrane lipid composition. Nat Rev Mol Cell Biol 19, 281–296. https://doi.org/10.1038/nrm.2017.138

Hegde, R.S., Keenan, R.J., 2022. The mechanisms of integral membrane protein biogenesis. Nat Rev Mol Cell Biol 23, 107–124. https://doi.org/10.1038/s41580-021-00413-2

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., Heijne, G. von, 2005a. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433, 377–381. https://doi.org/10.1038/nature03216

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., von Heijne, G., 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450, 1026–1030.

Hessa, T., White, S.H., von Heijne, G., 2005b. Membrane insertion of a potassium-channel voltage sensor. Science 307, 1427. https://doi.org/10.1126/science.1109176

Hilton, K.L.F., Manwani, C., Boles, J.E., White, L.J., Ozturk, S., Garrett, M.D., Hiscock, J.R., 2021. The phospholipid membrane compositions of bacterial cells, cancer cell lines and biological samples from cancer patients. Chem Sci 12, 13273–13282. https://doi.org/10.1039/d1sc03597e

Hua, F., Li, K., Shang, S., Wang, F., Hu, Z., 2019. Immune Signaling and Autophagy Regulation. Adv Exp Med Biol 1206, 551–593. https://doi.org/10.1007/978-981-15-0602-4_26

Hullin-Matsuda, F., Taguchi, T., Greimel, P., Kobayashi, T., 2014. Lipid compartmentalization in the endosome system. Seminars in Cell & Developmental Biology, Endosome dynamics & Tubulogenesis 31, 48–56. https://doi.org/10.1016/j.semcdb.2014.04.010

Illergård, K., Kauko, A., Elofsson, A., 2011. Why are polar residues within the membrane core evolutionary conserved? Proteins 79, 79–91. https://doi.org/10.1002/prot.22859

Ismail, N., Crawshaw, S.G., Cross, B.C.S., Haagsma, A.C., High, S., 2008. Specific transmembrane segments are selectively delayed at the ER translocon during opsin biogenesis. Biochem J 411, 495–506. https://doi.org/10.1042/BJ20071597

Israelachvili, J.N., 1977. Refinement of the fluid-mosaic model of membrane structure. Biochim Biophys Acta 469, 221–225. https://doi.org/10.1016/0005-2736(77)90185-7

Jaud, S., Fernández-Vidal, M., Nilsson, I., Meindl-Beinker, N.M., Hübner, N.C., Tobias, D.J., von Heijne, G., White, S.H., 2009. Insertion of short transmembrane helices by the Sec61 translocon. Proc Natl Acad Sci U S A 106, 11588–11593. https://doi.org/10.1073/pnas.0900638106

Jomaa, A., Gamerdinger, M., Hsieh, H.-H., Wallisch, A., Chandrasekaran, V., Ulusoy, Z., Scaiola, A., Hegde, R.S., Shan, S.-O., Ban, N., Deuerling, E., 2022. Mechanism of signal sequence handover from NAC to SRP on ribosomes during ER-protein targeting. Science 375, 839–844. https://doi.org/10.1126/science.abl6459

Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., Schuldiner, M., 2009. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Science 323, 1693–1697. https://doi.org/10.1126/science.1167983

Kanada, S., Takeguchi, Y., Murakami, M., Ihara, K., Kouyama, T., 2011. Crystal Structures of an O-Like Blue Form and an Anion-Free Yellow Form of pharaonis Halorhodopsin. Journal of Molecular Biology 413, 162–176. https://doi.org/10.1016/j.jmb.2011.08.021

Kanonenberg, K., Royes, J., Kedrov, A., Poschmann, G., Angius, F., Solgadi, A., Spitz, O., Kleinschrodt, D., Stühler, K., Miroux, B., Schmitt, L., 2019. Shaping the lipid composition of bacterial membranes for membrane protein production. Microbial Cell Factories 18, 131. https://doi.org/10.1186/s12934-019-1182-1

Kelekar, A., Chang, B.S., Harlan, J.E., Fesik, S.W., Thompson, C.B., 1997. Bad is a BH3 domain-containing protein that forms an inactivating dimer with Bcl-XL. Mol Cell Biol 17, 7040–7046.

Kerppola, T.K., 2006. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. Nat Protoc 1, 1278–1286. https://doi.org/10.1038/nprot.2006.201

Kerr, J.F., Wyllie, A.H., Currie, A.R., 1972. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 26, 239–257. https://doi.org/10.1038/bjc.1972.33

Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D.L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D.S., Fleishman, S.J., 2018. Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. Mol. Cell 72, 178-186.e5. https://doi.org/10.1016/j.molcel.2018.08.033

Kim, H., Rafiuddin-Shah, M., Tu, H.-C., Jeffers, J.R., Zambetti, G.P., Hsieh, J.J.-D., Cheng, E.H.-Y., 2006. Hierarchical regulation of mitochondrion-dependent apoptosis by BCL-2 subfamilies. Nat Cell Biol 8, 1348–1358. https://doi.org/10.1038/ncb1499

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kvansakul, M., Caria, S., Hinds, M.G., 2017. The Bcl-2 Family in Host-Virus Interactions. Viruses 9, 290. https://doi.org/10.3390/v9100290

Kvansakul, M., Hinds, M.G., 2013. Structural biology of the Bcl-2 family and its mimicry by viral proteins. Cell Death Dis 4, e909. https://doi.org/10.1038/cddis.2013.436

Lee, H., Kim, H., 2014. Membrane topology of transmembrane proteins: determinants and experimental tools. Biochem Biophys Res Commun 453, 268–276. https://doi.org/10.1016/j.bbrc.2014.05.111

Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., Engelman, D.M., 1992a. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. J. Biol. Chem. 267, 7683–7689.

Lemmon, M.A., Flanagan, J.M., Treutlein, H.R., Zhang, J., Engelman, D.M., 1992b. Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry 31, 12719–12725. https://doi.org/10.1021/bi00166a002

Lin, C.-Y., Lin, L.-Y., 2018. The conserved basic residues and the charged amino acid residues at the α-helix of the zinc finger motif regulate the nuclear transport activity of triple C2H2 zinc finger proteins. PLOS ONE 13, e0191971. https://doi.org/10.1371/journal.pone.0191971

Lucendo, E., Sancho, M., Lolicato, F., Javanainen, M., Kulig, W., Leiva, D., Duart, G., Andreu-Fernández, V., Mingarro, I., Orzáez, M., 2020. Mcl-1 and Bok transmembrane domains: Unexpected players in the modulation of apoptosis. PNAS. https://doi.org/10.1073/pnas.2008885117

Lucianò, A.M., Pérez-Oliva, A.B., Mulero, V., Del Bufalo, D., 2021. Bcl-xL: A Focus on Melanoma Pathobiology. International Journal of Molecular Sciences 22, 2777. https://doi.org/10.3390/ijms22052777

Lučić, V., Leis, A., Baumeister, W., 2008. Cryo-electron tomography of cells: connecting structure and function. Histochem Cell Biol 130, 185–196. https://doi.org/10.1007/s00418-008-0459-y

MacCallum, J.L., Tieleman, D.P., 2011. Hydrophobicity scales: a thermodynamic looking glass into lipid–protein interactions. Trends in Biochemical Sciences 36, 653–662. https://doi.org/10.1016/j.tibs.2011.08.003

Marqusee, S., Baldwin, R.L., 1987. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. Proc. Natl. Acad. Sci. U. S. A. 84, 8898–8902.

Martin, J., Sawyer, A., 2019. Elucidating the structure of membrane proteins. BioTechniques 66, 167–170. https://doi.org/10.2144/btn-2019-0030

Martinez-Gil, L., Mingarro, I., 2015. Viroporins, Examples of the Two-Stage Membrane Protein Folding Model. Viruses 7, 3462–3482. https://doi.org/10.3390/v7072781

Martínez-Gil, L., Saurí, A., Marti-Renom, M.A., Mingarro, I., 2011. Membrane protein integration into the endoplasmic reticulum. FEBS J 278, 3846–3858. https://doi.org/10.1111/j.1742-4658.2011.08185.x

McGilvray, P.T., Anghel, S.A., Sundaram, A., Zhong, F., Trnka, M.J., Fuller, J.R., Hu, H., Burlingame, A.L., Keenan, R.J., 2020. An ER translocon for multi-pass membrane protein biogenesis. Elife 9, e56889. https://doi.org/10.7554/eLife.56889

Meacock, S.L., Lecomte, F.J.L., Crawshaw, S.G., High, S., 2002. Different transmembrane domains associate with distinct endoplasmic reticulum components during membrane integration of a polytopic protein. Mol Biol Cell 13, 4114–4129. https://doi.org/10.1091/mbc.e02-04-0198

Mingarro, I., von Heijne, G., Whitley, P., 1997. Membrane-protein engineering. Trends in Biotechnology 15, 432–437. https://doi.org/10.1016/S0167-7799(97)01101-3

Moore, D.T., Berger, B.W., DeGrado, W.F., 2008. Protein-protein interactions in the membrane: sequence, structural, and biological motifs. Structure 16, 991–1001. https://doi.org/10.1016/j.str.2008.05.007

Murzyn, K., Róg, T., Pasenkiewicz-Gierula, M., 2005. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. Biophys J 88, 1091–1103. https://doi.org/10.1529/biophysj.104.048835

Nichols, D.B., De Martini, W., Cottrell, J., 2017. Poxviruses Utilize Multiple Strategies to Inhibit Apoptosis. Viruses 9, 215. https://doi.org/10.3390/v9080215

Nicolson, G.L., 2014. The Fluid-Mosaic Model of Membrane Structure: still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. Biochim Biophys Acta 1838, 1451–1466. https://doi.org/10.1016/j.bbamem.2013.10.019

Nilsson, I., von Heijne, G., 1990. Fine-tuning the topology of a polytopic membrane protein: Role of positively and negatively charged amino acids. Cell 62, 1135–1141. https://doi.org/10.1016/0092-8674(90)90390-Z

Nilsson, J., Persson, B., von Heijne, G., 2005. Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. Proteins 60, 606–616. https://doi.org/10.1002/prot.20583

Oberai, A., Ihm, Y., Kim, S., Bowie, J.U., 2006. A limited universe of membrane protein families and folds. Protein Science 15, 1723–1734. https://doi.org/10.1110/ps.062109706

Oltvai, Z.N., Milliman, C.L., Korsmeyer, S.J., 1993. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. Cell 74, 609–619. https://doi.org/10.1016/0092-8674(93)90509-o

O'Neill, J.W., Manion, M.K., Maguire, B., Hockenbery, D.M., 2006. BCL-XL dimerization by three-dimensional domain swapping. J Mol Biol 356, 367–381. https://doi.org/10.1016/j.jmb.2005.11.032

Overington, J.P., Al-Lazikani, B., Hopkins, A.L., 2006. How many drug targets are there? Nat Rev Drug Discov 5, 993–996. https://doi.org/10.1038/nrd2199

Pearson, G.R., Luka, J., Petti, L., Sample, J., Birkenbach, M., Braun, D., Kieff, E., 1987. Identification of an Epstein-Barr virus early gene encoding a second component of the restricted early antigen complex. Virology 160, 151–161. https://doi.org/10.1016/0042-6822(87)90055-9

Peschke, M., Le Goff, M., Koningstein, G.M., Karyolaimos, A., de Gier, J.-W., van Ulsen, P., Luirink, J., 2018. SRP, FtsY, DnaK and YidC Are Required for the Biogenesis of the E. coli Tail-Anchored Membrane Proteins DjlC and Flk. J Mol Biol 430, 389–403. https://doi.org/10.1016/j.jmb.2017.12.004

Polčic, P., Mentel, M., Gavurníková, G., Bhatia-Kiššová, I., 2017. To keep the host alive - the role of viral Bcl-2 proteins. Acta Virol 61, 240–251. https://doi.org/10.4149/av_2017_302

Popot, J.-L., Engelman, D.M., 2000. Helical Membrane Protein Folding, Stability, and Evolution. Annual Review of Biochemistry 69, 881–922. https://doi.org/10.1146/annurev.biochem.69.1.881

Popot, J.L., Engelman, D.M., 1990. Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29, 4031–4037. https://doi.org/10.1021/bi00469a001

Rooswinkel, R.W., van de Kooij, B., de Vries, E., Paauwe, M., Braster, R., Verheij, M., Borst, J., 2014. Antiapoptotic potency of Bcl-2 proteins primarily relies on their stability, not binding selectivity. Blood 123, 2806–2815. https://doi.org/10.1182/blood-2013-08-519470

Rotello, R.J., Veenstra, T.D., 2021. Mass Spectrometry Techniques: Principles and Practices for Quantitative Proteomics. Curr Protein Pept Sci 22, 121–133. https://doi.org/10.2174/1389203721666200921153513

Ruan, W., Lindner, E., Langosch, D., 2004. The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis. Protein Sci 13, 555–559. https://doi.org/10.1110/ps.03357404

Ruch, T.R., Machamer, C.E., 2012. The Coronavirus E Protein: Assembly and Beyond. Viruses 4, 363–382. https://doi.org/10.3390/v4030363

Sanderson, J.M., 2012. Resolving the kinetics of lipid, protein and peptide diffusion in membranes. Molecular Membrane Biology 29, 118–143. https://doi.org/10.3109/09687688.2012.678018

Schanzenbach, C., Schmidt, F.C., Breckner, P., Teese, M.G., Langosch, D., 2017. Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions. Scientific Reports 7, 43476. https://doi.org/10.1038/srep43476

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., Weissman, J.S., 2008. The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell 134, 634–645. https://doi.org/10.1016/j.cell.2008.06.025

Shurtleff, M.J., Itzhak, D.N., Hussmann, J.A., Schirle Oakdale, N.T., Costa, E.A., Jonikas, M., Weibezahn, J., Popova, K.D., Jan, C.H., Sinitcyn, P., Vembar, S.S., Hernandez, H., Cox, J., Burlingame, A.L., Brodsky, J.L., Frost, A., Borner, G.H., Weissman, J.S., 2018. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. eLife 7, e37018. https://doi.org/10.7554/eLife.37018

Singer, S.J., 1974. The molecular organization of membranes. Annu Rev Biochem 43, 805–833. https://doi.org/10.1146/annurev.bi.43.070174.004105

Singer, S.J., Nicolson, G.L., 1972. The Fluid Mosaic Model of the Structure of Cell Membranes. Science 175, 720–731. https://doi.org/10.1126/science.175.4023.720

Sohlenkamp, C., Geiger, O., 2016. Bacterial membrane lipids: diversity in structures and pathways. FEMS Microbiology Reviews 40, 133–159. https://doi.org/10.1093/femsre/fuv008

Spiess, M., Lodish, H.F., 1986. An internal signal sequence: the asialoglycoprotein receptor membrane anchor. Cell 44, 177–185. https://doi.org/10.1016/0092-8674(86)90496-4

Stillwell, W., 2013. Chapter 6 - Membrane Proteins, in: Stillwell, W. (Ed.), An Introduction to Biological Membranes. Elsevier, San Diego, pp. 85–105. https://doi.org/10.1016/B978-0-444-52153-8.00006-4

Stoeckenius, W., Engelman, D.M., 1969. CURRENT MODELS FOR THE STRUCTURE OF BIOLOGICAL MEMBRANES. J Cell Biol 42, 613–646.

Subczynski, W.K., Pasenkiewicz-Gierula, M., Widomska, J., Mainali, L., Raguz, M., 2017. High cholesterol/low cholesterol: Effects in biological membranes Review. Cell Biochem Biophys 75, 369–385. https://doi.org/10.1007/s12013-017-0792-7

Tamborero, S., Vilar, M., Martínez-Gil, L., Johnson, A.E., Mingarro, I., 2011. Membrane Insertion and Topology of the Translocating Chain-Associating Membrane Protein (TRAM). Journal of Molecular Biology 406, 571–582. https://doi.org/10.1016/j.jmb.2011.01.009

Toyoshima, C., Nakasako, M., Nomura, H., Ogawa, H., 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. Nature 405, 647–655. https://doi.org/10.1038/35015017

Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A.-K., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.-H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J., Pontén, F., 2015. Proteomics. Tissue-based map of the human proteome. Science 347, 1260419. https://doi.org/10.1126/science.1260419

Van den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., Rapoport, T.A., 2004. X-ray structure of a protein-conducting channel. Nature 427, 36–44. https://doi.org/10.1038/nature02218

van Meer, G., Voelker, D.R., Feigenson, G.W., 2008. Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9, 112–124. https://doi.org/10.1038/nrm2330

Verdiá-Báguena, C., Nieto-Torres, J.L., Alcaraz, A., DeDiego, M.L., Torres, J., Aguilella, V.M., Enjuanes, L., 2012. Coronavirus E protein forms ion channels with functionally and structurally-involved membrane lipids. Virology 432, 485–494. https://doi.org/10.1016/j.virol.2012.07.005

Vereb, G., Szöllosi, J., Matkó, J., Nagy, P., Farkas, T., Vigh, L., Mátyus, L., Waldmann, T.A., Damjanovich, S., 2003. Dynamic, yet structured: The cell membrane three decades after the Singer-Nicolson model. Proc Natl

Acad Sci U S A 100, 8053–8058. https://doi.org/10.1073/pnas.1332550100

von Heijne, G., 2006. Membrane-protein topology. Nat Rev Mol Cell Biol 7, 909–918. https://doi.org/10.1038/nrm2063

von Heijne, G., 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 341, 456–458.

von Heijne, G., 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. EMBO J 5, 3021–3027.

Voorhees, R.M., Hegde, R.S., 2015. Structures of the scanning and engaged states of the mammalian SRP-ribosome complex. eLife 4, e07975. https://doi.org/10.7554/eLife.07975

Walters, R.F.S., DeGrado, W.F., 2006. Helix-packing motifs in membrane proteins. Proc Natl Acad Sci U S A 103, 13658–13663. https://doi.org/10.1073/pnas.0605878103

Walther, T.H., Ulrich, A.S., 2014. Transmembrane helix assembly and the role of salt bridges. Current Opinion in Structural Biology, Membranes / Engineering and design 27, 63–68. https://doi.org/10.1016/j.sbi.2014.05.003

Wang, K., Yin, X.M., Chao, D.T., Milliman, C.L., Korsmeyer, S.J., 1996. BID: a novel BH3 domain-only death agonist. Genes Dev 10, 2859–2869. https://doi.org/10.1101/gad.10.22.2859

Wasilenko, S.T., Stewart, T.L., Meyers, A.F.A., Barry, M., 2003. Vaccinia virus encodes a previously uncharacterized mitochondrial-associated inhibitor of apoptosis. Proc Natl Acad Sci U S A 100, 14345–14350. https://doi.org/10.1073/pnas.2235583100

Weinstein, J.Y., Elazar, A., Fleishman, S.J., 2019. A lipophilicity-based energy function for membrane-protein modelling and design. PLOS Computational Biology 15, e1007318. https://doi.org/10.1371/journal.pcbi.1007318

Westphal, D., Ledgerwood, E.C., Hibma, M.H., Fleming, S.B., Whelan, E.M., Mercer, A.A., 2007. A novel Bcl-2-like inhibitor of apoptosis is encoded by the parapoxvirus ORF virus. J Virol 81, 7178–7188. https://doi.org/10.1128/JVI.00404-07

White, S.H., Wimley, W.C., 1999. Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 28, 319–365. https://doi.org/10.1146/annurev.biophys.28.1.319

Whitley, P., Grau, B., Gumbart, J.C., Martínez-Gil, L., Mingarro, I., 2021. Folding and Insertion of Transmembrane Helices at the ER. Int J Mol Sci 22, 12778. https://doi.org/10.3390/ijms222312778

Whitley, P., Mingarro, I., 2014. Stitching proteins into membranes, not sew simple. Biol Chem 395, 1417–1424. https://doi.org/10.1515/hsz-2014-0205

Wiener, M.C., White, S.H., 1992. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. II. Distribution and packing of terminal methyl groups. Biophys J 61, 428–433.

Wimley, W.C., Creamer, T.P., White, S.H., 1996. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host−Guest Pentapeptides. Biochemistry 35, 5109–5124. https://doi.org/10.1021/bi9600153

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., Meng, J., Zhu, Z., Zhang, Z., Wang, J., Sheng, J., Quan, L., Xia, Z., Tan, W., Cheng, G., Jiang, T., 2020. Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host & Microbe 27, 325–328. https://doi.org/10.1016/j.chom.2020.02.001

Xia, B., Shen, X., He, Y., Pan, X., Liu, F.-L., Wang, Y., Yang, F., Fang, S., Wu, Y., Duan, Z., Zuo, X., Xie, Z., Jiang, X., Xu, L., Chi, H., Li, S., Meng, Q., Zhou, H., Zhou, Y., Cheng, X., Xin, X., Jin, L., Zhang, H.-L., Yu, D.-D., Li, M.-H., Feng, X.-L., Chen, J., Jiang, H., Xiao, G., Zheng, Y.-T., Zhang, L.-K., Shen, J., Li, J., Gao, Z., 2021. SARS-CoV-2 envelope protein causes acute respiratory distress syndrome (ARDS)-like pathological damages and constitutes an antiviral target. Cell Res 31, 847–860. https://doi.org/10.1038/s41422-021-00519-4

Xie, Z., Schendel, S., Matsuyama, S., Reed, J.C., 1998. Acidic pH promotes dimerization of Bcl-2 family proteins. Biochemistry 37, 6410–6418. https://doi.org/10.1021/bi973052i

Yao, X., Fan, X., Yan, N., 2020. Cryo-EM analysis of a membrane protein embedded in the liposome. PNAS 117, 18497–18503. https://doi.org/10.1073/pnas.2009385117

Yeung, T., Gilbert, G.E., Shi, J., Silvius, J., Kapus, A., Grinstein, S., 2008. Membrane phosphatidylserine regulates surface charge and protein localization. Science 319, 210–213. https://doi.org/10.1126/science.1152066

Yin, H., Slusky, J.S., Berger, B.W., Walters, R.S., Vilaire, G., Litvinov, R.I., Lear, J.D., Caputo, G.A., Bennett, J.S., DeGrado, W.F., 2007. Computational Design of Peptides That Target Transmembrane Helices. Science 315, 1817–1822. https://doi.org/10.1126/science.1136782

Zamzami, N., Brenner, C., Marzo, I., Susin, S.A., Kroemer, G., 1998. Subcellular and submitochondrial mode of action of Bcl-2-like oncoproteins. Oncogene 16, 2265–2282. https://doi.org/10.1038/sj.onc.1201989

Zhang, F., Ge, W., Ruan, G., Cai, X., Guo, T., 2020. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. Proteomics 20, e1900276. https://doi.org/10.1002/pmic.201900276

Zhang, X., Shan, S., 2014. Fidelity of Co-translational Protein Targeting by the Signal Recognition Particle. Annu Rev Biophys 43, 381–408. https://doi.org/10.1146/annurev-biophys-051013-022653

Zheng, M., Karki, R., Williams, E.P., Yang, D., Fitzpatrick, E., Vogel, P., Jonsson, C.B., Kanneganti, T.-D., 2021. TLR2 senses the SARS-CoV-2 envelope protein to produce inflammatory cytokines. Nat Immunol 22, 829–838. https://doi.org/10.1038/s41590-021-00937-x

Zhou, F.X., Cocco, M.J., Russ, W.P., Brunger, A.T., Engelman, D.M., 2000. Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat Struct Biol 7, 154–160. https://doi.org/10.1038/72430

8. Annexes

8.1. Annex I

Duart, G., Grau, B., Mingarro, I., Martinez-Gil, L., 2021b. Methodological approaches for the analysis of transmembrane domain interactions: A systematic review. *Biochim Biophys Acta Biomembr* 1863, 183712. https://doi.org/10.1016/j.bbamem.2021.183712

Review

Methodological approaches for the analysis of transmembrane domain interactions: A systematic review

Gerard Duart, Brayan Grau, Ismael Mingarro, Luis Martinez-Gil [*]

Department of Biochemistry and Molecular Biology, Institute for Biotechnology and Biomedicine BioTecMed, University of Valencia, Dr. Moliner 50, 46100 Burjassot, Spain

ABSTRACT

The study of protein-protein interactions (PPI) has proven fundamental for the understanding of the most relevant cell processes. Any protein domain can participate in PPI, including transmembrane (TM) segments that can establish interactions with other TM domains (TMDs). However, the hydrophobic nature of TMDs and the environment they occupy complicates the study of intramembrane PPI, which demands the use of specific approaches and techniques. In this review, we will explore some of the strategies available to study intramembrane PPI *in vitro*, *in vivo*, and, *in silico*, focusing on those techniques that could be carried out in a standard molecular biology laboratory regarding its previous experience with membrane proteins.

1. Introduction

The study of protein-protein interactions (PPI) has proven fundamental for the understanding of any cell process. Over the last decade, with the *omic* upsurge and the vast increase in computing power, our knowledge of PPI networks and their implications over the cell functions has increased vastly, *e.g.* according to the string database over 2000 million interactions have been described for 5090 organisms [1]. The interest in PPI has not been limited to the big picture, the fine details that govern these interactions have also received a lot of attention, particularly, from the pharmaceutical industry. Understanding the minutiae of PPI and protein-small-molecule interactions is critical for rational drug design [3]. Additionally, virtual screenings, based on structural and protein contact information, have become a common strategy in the identification and optimization of lead compounds [4,5]. In this quest for understanding how proteins interact there is still one big gap, and that is, the PPI occurring within the membrane milieu. Membrane proteins (MPs), those embedded in biological membranes, represent roughly 25% of the human proteome and almost 60% of the current drugs are designed towards MPs [6]. Recent work has sparked some interest in their role as PPI domains and thus as potential drug targets. The hydrophobic nature of TMDs, and the environment which they occupy, complicates the biochemical study of intramembrane PPI. These difficulties in handling MPs demand the use of specific approaches and techniques. In this review, we will explore some of the strategies available to study intramembrane PPI, focusing on those techniques that could be carried out in a standard molecular biology laboratory regarding its previous experience with membrane proteins. Our review aims to demystify the work with TMDs so more comprehensive studies are carried out with MPs. The methodologies included in this review were designed/adapted to study TMD interactions. That is, no other (extra-membranous) parts of the proteins are included in the assays that can influence the observed results.

2. Monitoring TMD oligomerization with the ToxR transcription activator

A methodology based on the dimerization of the ToxR transcription activator was the first user-friendly approach developed for the study of TMD homo-oligomerization [7,8]. This methodology used a chimeric construct in which the cytoplasmic DNA binding domain of ToxR, a dimerization-dependent transcriptional activator, is fused to a TMD of interest and a monomeric periplasmic located protein (in this case, the maltose-binding protein, MBP) (Fig. 1). The resulting chimera is expressed in *E. coli* where self-association of the TMD results in the ToxR-mediated expression of the β-galactosidase as a reporter enzyme. The system allows a quantitative analysis of the interaction since the level of β-galactosidase activity correlates with the strength of TMDs association.

Incorporation of the MBP in the chimeric construct ensures its

* Corresponding author.
 E-mail address: luis.martinez-gil@uv.es (L. Martinez-Gil).

correct topology in the bacterial inner membrane, minimizing the possibility of non-specific hydrophobic interactions in the cell cytosol. To this end, the ToxR-based assay is performed in a bacteria strain lacking the endogenous MBP (malE-deficient). Consequently, these cells cannot transport maltose into the cytoplasm and grow when the only available source of carbon is maltose. Therefore, only when the chimeric construct is expressed and inserted in the bacteria membrane with the correct orientation, that is, with the ToxR domain facing the cytosol and the MBP in the periplasm (Fig. 1), the latter will complement the *malE*-deficient phenotype.

A few years after the publication of the original assay the ß-galactosidase in the chimera was replaced by the chloramphenicol acetyltransferase (CAT) as a reporter. Activation of the CAT provides antibiotic (chloramphenicol) resistance and facilitates the growth of the bacteria in selective media, this variation of the original design was named ToxCAT [9]. Later on, the ToxCAT assay was utilized to isolate previously unidentified TMD dimerization motifs from a library of randomized TMDs [10]. In this assay, randomization was limited to seven positions within a 19 residue-long stretch, while the remaining positions were held constant as either alanines or leucines. The seven positions that were allowed to change (1, 2, 5, 6, 9, 10, and 13 respectively) define the interface of a hypothetical right-handed alpha-helix TMD dimer, a design based on the structure of the GpA homo-dimer [11]. Over 80% of the isolates positive for the ToxCAT assay contained a GxxxG motif, as GpA does. Furthermore, the adjacent flanking residues accompanying the GxxxG motif were size complementary at the potential dimer interface. Although the results were biased, due to the forced location of variable residues, these results identified GxxxG as a general sequence motif involved in the association of transmembrane helices. Nonetheless, the actual interaction strength of GxxxG-containing helices depends strongly on the sequence context and membrane properties [12]. In the context of the GxxxG motif, the energy for the association of TMDs is, mostly, driven by a combination of van der Waals interactions, which are maximized by the precise packing of complementary helix faces, as suggested previously by the limited structural information at the time [13] and hydrogen bonds between Cα-H carbon donors and carbonyl acceptors on opposing helices [14].

Since its publication, the ToxCAT approach has been used for the study of many TMD oligomers of prokaryotic and eukaryotic origin including those found in the *H. pylori* vacuolating cytotoxin [15], the ErbB receptor [16], or the integrin family [17] among others. Additionally, ToxCAT has helped to identify new interaction motifs [18] and to refine the basics of TMD association [19,20], particularly in the case of GpA dimer [21,22]. It is important to remember that, although the system admits eukaryotic TMDs the membrane environment in which those are placed might not accurately resemble the native location of these segments. These differences in hydrophobic length, lipid composition, or charge balance might impact the propensity of the TMDs to interact.

During all these years, the system has been modified in several ways to overcome some of its shortcomings and to accommodate the necessities of multiple laboratories. Early in 2006, a modified dual reporter ToxCAT variation was published [23]. In this modification of the original methodology, the MBP was substituted by the β-lactamase, allowing simultaneous monitoring of the chimeric protein insertion in the membrane (through the exposure of the β-lactamase to the periplasm and the antibiotic resistance that it confers) and the TMD ability to homo-oligomerize (through the CAT reporter). The authors claimed that this approach is particularly well suited for the identification of random mutations in the TMD which, might not only affect TMD oligomerization, but also TMD targeting and insertion.

This bacterial system has also been adapted for the study of heterodimers [24]. The system known as GALLEX replaces the ToxR domain with the LexA DNA binding domain fused to a TMD of interest and the MBP in the C-terminal end. The LexA DNA binding domain variants bind to different DNA sequences thus enabling the analysis of TMD heterooligomerizations [25]. Briefly, two LexA DNA binding domains with different DNA sequence specificity are coupled to the assayed TMDs. Interaction of the TMD leads to the formation of LexA heterodimers, which can bind to a promoter/operator sequence containing one specific binding site for each LexA variant. Binding of the LexA heterodimer results in repression of the *lacZ* gene, which serves as an indication of TMD interaction.

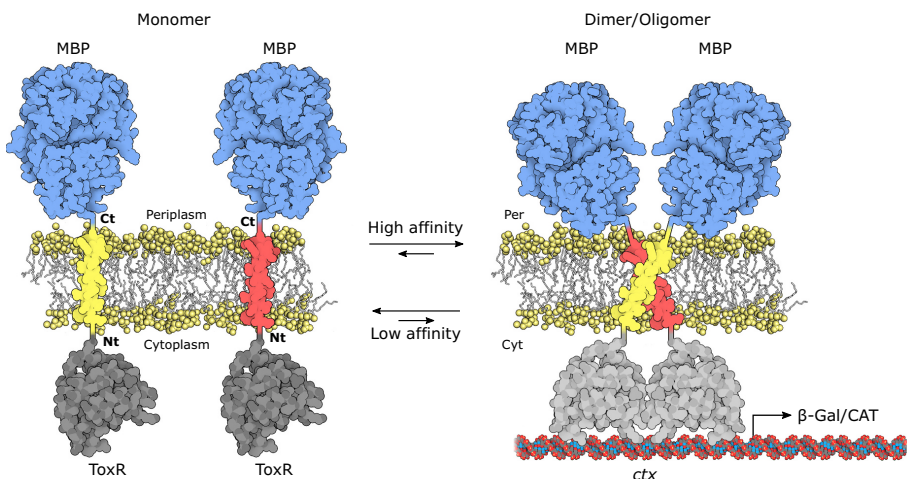Steindorf et al. inspired by the BACTH [26] system, generated

Fig. 1. Schematic representation of the ToxR-based assay. TMD-mediated oligomerization of ToxR (grey) results in β-gal or CAT activation at the *ctx* promoter. The periplasmic C-terminal MBP domain (blue) ensures the selection of chimeras anchored to the inner membrane of *E. coli*. The position of the N-terminus and C-terminus ends and the TMDs are indicated. Protein representations were done with *Illustrate* [105]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

another variation of this bacterial assay to identify TMD hetero-dimers [27]. In this variation, also conducted in an *E. coli* strain lacking the endogenous MBP, the tested TMDs are N-terminally fused to either the T18 or T25-fragment of the *Bordetella pertussis* adenylate cyclase [26], and C-terminally fused to the *malE* gene. Interaction of the two chimeric proteins results in the reconstitution of the adenylate cyclase and thereby restores its enzymatic activity. The, resulting increase in cAMP levels activates the *E. coli mal* regulon, facilitates cell growth of a *malE*-deficient strain on media containing maltose as the only carbohydrate source.

Berger et al. also re-design the ToxR-based system to analyze intra-membrane hetero-oligomerizations [28]. In this case, two chimeric constructs are co-expressed in *E. coli*. While one of them presents the TMD fused to the ToxR transcriptional activator (as in the original design) and can activate transcription at the *ctx* promoter upon dimerization; the other has the TMD fused to an inactive ToxR mutant (ToxR*). Upon the formation of a TMD-TMD hetero-oligomer, the inactive ToxR* exerts a dominant-negative effect on the reporter gene synthesis. This modification was accompanied by the substitution of the reporter by the Red Fluorescent Protein (RFP), which facilitated the signal quantification. Other groups have also substituted the β-galactosidase or CAT reporters with fluorescent proteins, *e.g.* the super-folded Green Fluorescence Protein (GFP) [29]. This, so-called ToxGREEN assay, was developed to facilitate the detection of fluorescence directly in unprocessed cell cultures and, according to the authors, significantly reduces the time, labor and, cost associated with the procedure, simplifying the screening of a large number of TMDs simultaneously using 96 well plates. Alternatively, the ToxR-based assays have included other reporter proteins such as the luciferase [30].

Not only were the ToxR-based approaches among the firsts methodologies for the analysis of TMD-TMD interactions, but it is a simple, powerful, and versatile technique. In consequence, it has been widely used in those labs exploring intramembrane interactions in the past two decades. Furthermore, it has become a primer in the design and implementation of other techniques for the study of TMD homo- and hetero-oligomerizations.

3. BiFC, Bimolecular Fluorescent Complementation assay

One of the most widely used approaches for the study of protein-protein interactions (PPI) is the bimolecular fluorescence complementation (BiFC) assay. The BiFC approach was designed, nearly two decades ago by Kerppola and colleagues, to obtain an *in situ* quantitative and visual PPI assay that did not require large instrumentation, as was not the case with other fluorescent approaches [31]. The assay utilizes the structural properties of the GFP and its derivates, a fluorescent protein family that can withstand peptide insertions, protein fusion, and protein splitting while retaining its fluorescence [32–34]. The BIFC approach requires splitting a fluorescent protein into two fragments, none of which must exhibit detectable fluorescence or show any propensity for interaction with the complementary fragment. In a seminal work, the GFP was split and each fragment fused to one of the long alpha-helices of a parallel leucine zipper [33]. The resulting chimeric constructs were expressed in *E. coli* where the leucine zipper could retain its capacity to interact, thus bringing nearby both fragments of the split GFP. Having both fragments of the split GFP close enough allowed for the reconstitution of its structure and its fluorescence properties. The authors quickly realized and proposed, that fragmented GFP reconstitution could be used for PPI studies in cells. Indeed, since its implementation, BiFC has been used in thousands of publications in all type of model organisms/systems including *E. coli* [33,35], yeast [36,37], cultured plant cells [38], cultured mammalian cells [39,40], plants [41,42], and even in mouse [43].

Once the interacting partners find each other, the bimolecular fluorescent complex is formed and the fluorophore is reconstituted. Interestingly, due to the structural stability of the reconstituted fluorescent protein, approximately 1 min after the interaction partners form a complex, isomerization takes place in the newly reconstituted fluorescent protein that leads to a stable association (reconstitution $t_{1/2} \sim 60$ s, depending on the conditions where the process occurs [31]). Fluorescent detection measurements should be delayed up to 1 h after fluorophore reconstitution to allow full complex maturation (complex maturation $t_{1/2} \sim 3000$ s) [31].

This, apparently simple assay, has many advantages for studying PPIs. The design of a BiFC assay does not require any structural information of the potential interacting partners. Additionally, the fluorescent nature of the assay allows for the visualization of the cellular location where the interaction takes place under physiological conditions. Furthermore, little disturbance is applied to the cells due to the absence of exogenous enzymatic activities or substrates. Another advantage of this methodology is the stability of the reconstituted fluorophore which provides the assay with the ability to detect weak and transient interactions without interference from non-specific interactions. The assay also offers the possibility of performing a multicolor analysis, where different fluorescent protein fragments can be fused to multiple interaction partners, allowing visualization of several PPI simultaneously using different excitation and emission wavelengths [31]. This variation of the original assay provides a novel way to study protein interaction networks.

The BiFC assay permits the identification and analysis of PPI within the membrane (Fig. 2) [40,44–46]. Nonetheless, the environment in which these interactions occur and the nature of membrane proteins requires the introduction of some modifications. First, it is important to remember that when studying TMD-TMD interactions the topology of the TMD in its functional context should be maintained. Therefore, the fluorescent protein halves should be fused either to the N- or the C-terminal end of the TMD ensuring the cytoplasmatic localization of the fluorescent protein halves and the TMD orientation required for the interaction [47]. In some cases, the addition of a signal peptide is necessary to target the chimera to the appropriated organelle and to achieve the desired topology. The BiFC approach does neither report on the insertion capacity nor the membrane topology of the studied TMDs as some other approaches do (see Sections 2 and 4). Therefore, other assays must be set alongside BiFC to ensure proper insertion into the membrane of the hydrophobic sequences under study and their topology. Bear in mind that, even when these conditions have been met, a fraction of the chimeric proteins may not insert or fold properly, which can lead to spurious PPIs out from the membrane.

It has been demonstrated that fluorescent proteins can be split at different points while retaining their reconstitution properties [32]. Nonetheless, an appropriate selection of the split point is important for BiFC efficiency since not all of them provide the same fluorescent characteristics to the assay, *i.e.*, different fluorescent spectra or maturation properties [48]. For the study of TMD oligomerization, the Venus fluorescent protein (VFP, a.k.a. SEYPF-F46L), a variant of the widely known Enhanced Yellow Fluorescent Protein (EYFP), has been commonly the fluorescent protein of choice. In addition to the EYFP properties the amino acid substitutions that shape the VFP (F46L, F64L, M153T, V163A, and, S175G), facilitate chromophore maturation by enhancing the protein folding at 37 °C and removing pre-incubation at lower temperatures before fluorescence measurements [49]. Furthermore, thanks to the properties of the VFP, the experiments can be performed with low protein amounts and reduced expression times, which minimize nonspecific fluorescence because of chimeras overcrowding in the membrane and the limited freedom of a 2D system such as the lipid bilayer [39]. For TMD oligomerization studies the N-terminal section of the VFP comprises residues 1–172 (VN172), while the C-terminal portion contains residues 155–238 (VC155), providing a 514/529 nm of maxima excitation and emission when reconstituted [49].

Due to the variability observed across experiments, it is recommended to normalize the fluorescence values obtained. Importantly, to avoid false positives generated by stochastic interactions between the
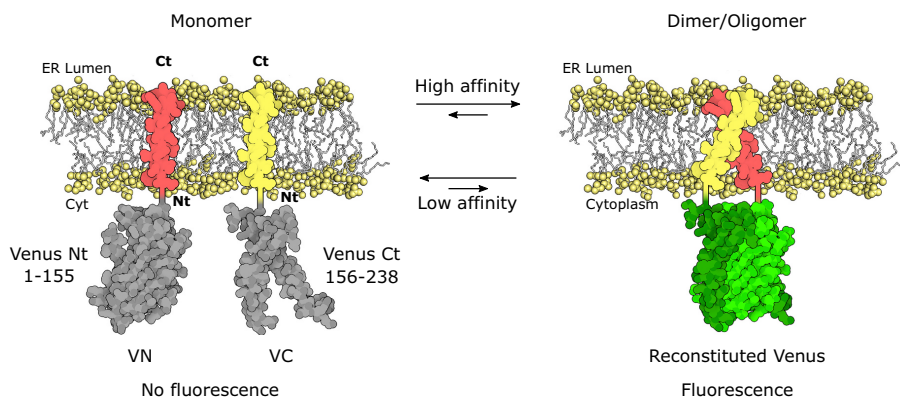
Fig. 2. Schematic representation of the Bimolecular Fluorescent Complementation (BiFC) assay. The two non-fluorescent fragments of the Venus fluorescent protein (Venus N-terminus (Nt) and Venus C-terminus (Ct); both in grey) are fused to two TMDs (yellow and red). The association of these TMDs allows the reconstitution of the fluorescent protein (green). The position of the N-terminus and C-terminus ends and the TMDs are indicated. The residues included in each VFP fragment are indicated below the protein representations. The periplasmic and cytoplasmic sides of the membrane are indicated as "Periplasm" (Per) and "Cytoplasm" (Cyt). Protein representations were done with *Illustrate* [105]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

two chimeric constructs, any given combination should be tested with appropriate negative controls [40,45]. It is recommended that the TMDs participating in the interaction under study are assayed in combination with a negative control *i.e.* a non-interacting TMD. As previously mentioned, for the study of TMD oligomerization one of the most used positive control, is that of GpA. Furthermore, few mutations in its dimerization motif (G79I and G83I) greatly reduce its association potential [50], rendering an excellent negative control [40,51].

In summary, the BiFC assay represents one of the most accessible approaches for the study of TMD homo- and hetero-oligomerizations due to its experimental simplicity. If appropriate controls are included, despite its experimental variability, the assay provides high-quality results for the identification of TMD oligomerization and the characterization of the amino acid residues responsible for this interaction. Furthermore, its versatility and wide range of applications, including drug discovery, make it a valuable tool in the repertoire of any protein research laboratory.

4. BLaTM, a split β-lactamase-based system for the analysis of interacting TMDs

The BLaTM assay is a recent addition to the toolbox for the study of homotypic and heterotypic interactions between TMDs in *E. coli* [52]. The system is based on the complementation of a split β-lactamase (Fig. 3). The β-lactamase used in the assay (TEM-1-β-lactamase) bears a mutation (M182T) [53] that increases the stability of the enzyme [54]. There are currently three different versions of the BLaTM system (v1.1, v1.2, and v2.0). For the initial BLaTM assay, the β-lactamase was split into two fragments, N-BLa (residues 23–164) and C-BLa (residues 196–286) [55], and fused, separately, to the Nt end of a TMD of interest *via* a flexible linker (containing G and S in an SGS(GGGS)$_2$GS sequence). Additionally, a super-folded green fluorescent protein (sfGFP) [54] was fused in-frame to the Ct end of the TMD *via* a rigid helical linker (containing A, K, and E residues in an A(EAAAK)$_5$A sequence) [55], and a FLAG-epitope is appended to the C-terminus of the sfGFP to facilitate chimeric protein detection by immunoblotting [52,56]. The chimeric proteins in BLaTM v1.1 and v1.2 contain an N-terminal cleavable signal peptide (SP) from pelbB (UniProtKB accession number Q00205, residues 1–20) or, ompA (UniProtKB accession number P0A910, residues 1–22) respectively, to ensure that both proteins are driven to the inner

bacterial membrane and inserted with an Nt-out topology (Fig. 3A) [52].

In all three versions of BLaTM, the resulting chimeric proteins (named after the fragment of the β-lactamase included, N-BLa and C-BLa respectively) are individually encoded in two low copy plasmids under the control of the pBAD arabinose promoter. While the N-BLa plasmid contains a p15A origin [57] and encodes CAT for chloramphenicol resistance, the C-BLa plasmid contains a pBR322 origin and encodes AraC and aminoglycoside 3′-phosphotransferase for kanamycin resistance [58]. This setup ensures that both plasmids are maintained within the same bacterial cell and provides control over protein expression through the addition of arabinose.

Once both chimeric proteins are properly inserted into the bacterial inner membrane, an interaction between their TMDs will facilitate the reconstitution of the β-lactamase structure providing the bacteria the means to grow in the presence of ampicillin [52]. Since β-lactamase activity requires its localization in the periplasmic space, the reconstituted β-lactamase will function both as an interaction and as a topological reporter.

In BLaTM, the efficiency of the TMD-TMD oligomerization is proportional to the reconstituted β-lactamase activity. Proportionality was elegantly demonstrated with BLaTM v1.1 through the study of two TMDs with different homo-oligomerizations propensities, the highly dimerizing TMD of GpA [9,59,60] and the medium-affinity TMD of quiescin sulfhydryl oxidase 2 (QSOX2) [60]. The observed LD$_{50}$ for ampicillin induced by the homo-dimerization of the GpA TMD was higher than the one induced by the QSOX2 TMD. Additionally, both homo-interactions and their associated LD$_{50}$ values were strongly diminished by point mutations known to disrupt the helix-helix interfaces (GpA G83I and QSOX2 S8A respectively) [9,60] confirming the specificity of the interaction and the correlation between oligomerization efficiency and the observed LD$_{50}$ values.

In addition to the arabinose control, the expression of pBAD promoter can be negatively regulated by isopropyl-thiogalactopyranoside (IPTG). Decreasing IPTG levels led to a progressive increase in the expression level accompanied by an increase in the associated LD$_{50}$ [52]. Apart from that, it was described that the expression of BLaTM v1.2 chimeric proteins tends to yield lower LD$_{50}$ values compared to BLaTM 1.1 due to the differences in the signal peptide used in each set up [52]. The lower expression level associated with BLaTM v1.2 together with the arabinose/IPTG control facilitates a fine-tune regulation of the TMD
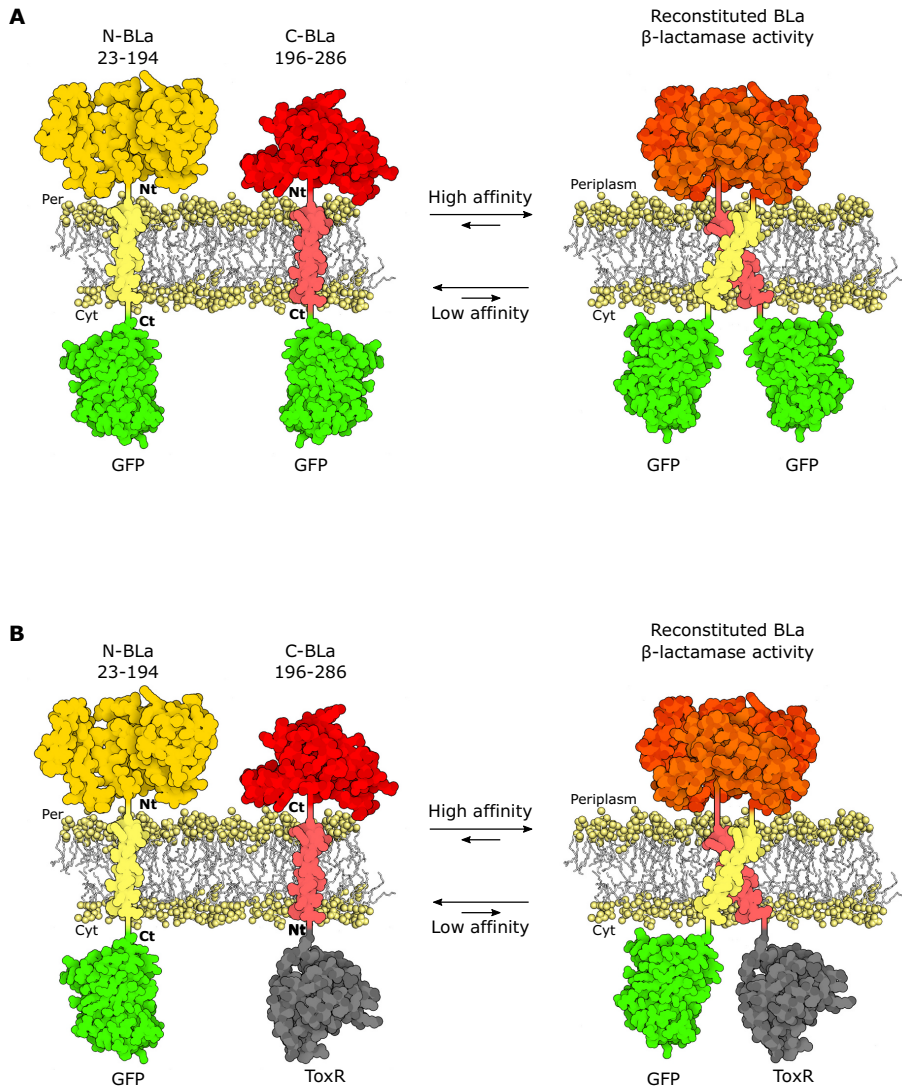
Fig. 3. Schematic representation of BLaTM assay. A) BlaTM v1.1 and 1.2. The two non-active fragments of the β-lactamase (N-BLa, bright yellow; C-Bla, bright red) are fused to two TMDs (N-terminus out; faded yellow and red) while the sfGFP (green) is fused to the TMD's C-terminus. The TMD oligomerization allows the reconstitution of the β-lactamase (orange) activity conferring ampicillin resistance only when the reconstituted β-lactamase is in the periplasmic side of the ER membrane. The position of the N- and C-terminus are indicated. B) BlaTM v2.0. The N-BLa non-active fragment of the β-lactamase is fused to a TMD in an N-terminus out orientation while the C-Bla non-active fragment is fused to a TMD in a C-terminus out orientation. As for v1.1 and 1.2, the N-BLa chimera is fused by the TMD to the sfGFP (green) in C-terminus. The C-Bla chimera is fused to a ToxR domain (grey) through the TMD N-terminus end to ensure the right topology. The TMD oligomerization allows the reconstitution of the β-lactamase activity conferring ampicillin resistance only when the reconstituted β-lactamase is in the periplasmic side of the ER membrane. The position of the N-terminus and C-terminus ends is indicated. The residues included in each β-lactamase fragment are indicated above the protein representations. The periplasmic and cytoplasmic sides of the membrane are indicated as "Periplasm" (Per) and "Cytoplasm" (Cyt). Protein representations were done with *Illustrate* [105]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

density in the bacterial membrane.

The BLaTM approach was used to show that the reporter signal resulting from a TMD-TMD interaction depends not only on the sequence of the interacting TMDs, and thus its inherent affinity, but also on the orientation of the TMDs interacting surfaces relative to the accompanying signaling domains [52,59,61]. To identify the optimal orientation of the tested TMDs, up to three amino acids were inserted at the TMD's N-terminal end and simultaneously removed at the C-terminus [52]. In this case, it was observed that the highest ampicillin resistance for model TMDs (GpA and QsOX2) was obtained when one or two amino acids respectively were inserted [52]. These results imply that an in-depth study of TMD-TMD interaction must be accompanied by the determination of the optimal orientation between the TMDs and the reporter signaling domains.

BLaTM v1.1 and v1.2 have been used to determine the strength of a TMD-TMD interaction through the determination of the LD_{50} in liquid media and to isolate high-affinity TMD-TMD partners from combinatorial libraries based on the differential TMD-TMD interaction-induced survival of bacteria on selective agar plates [52].

The experimental setup of BLaTM v1.1 and 1.2 requires that both interacting TMDs present the same topology in the membrane (parallel). More recently a third version of the system, BLaTM v2.0, was developed to study intramembrane interactions between oppositely orientated TMDs (antiparallel). Briefly, in BLaTM v2.0, the N-BLa chimeric protein described for BLaTM v1.2 is co-expressed with a C-BLa construct where the TMD is inserted in a reverse orientation (i.e., the C-BLa fragment being fused to the TMD's C-terminus). Additionally, in BLaTM v2.0 the sfGFP in C-BLa has been replaced by the cytoplasmic ToxR domain (see ToxR section in this manuscript), to ensure the N-terminus-in (cytoplasmatic) topology of a hybrid protein (Fig. 3B) [9,60].

BLaTM v2.0 was then used to study the homo-oligomerization of the TMD 4 of the dual-topology small multidrug transporter protein EmrE [62,63]. A comparison between BLaTM v1.2 and v2.0 showed that an anti-parallel interaction between two TMD 4 monomers is ~5-fold stronger than the parallel one. Nevertheless, it is worth mentioning that all tested TMD showing weak interactions in BLaTM v1.2 elicit higher LD_{50} values in BLaTM 2.0. This is probably due to the differences in the orientation of the BLa fragments affecting the efficiency of the β-lactamase reconstitution in both cases [64]. The setup in BLaTM v2.0 likely brings the Nt of the C-BLa fragment closer to the Ct of the N-BLa fragment in the N-BLa chimera. In this case, the authors proposed that the reconstitution of the β-lactamase occurs with a smaller loss of entropy than in the parallel configuration [64,65]. Therefore, to make a fair comparison between the LD_{50} obtained with a parallel and an antiparallel BLaTM setup it is recommended to use a TMD with expected similar low affinity as a reference point.

Interestingly, the fact that all required components in BLaTM (either in v1.1, v1.2, or v2.0) are encoded by plasmids, and no chromosomally integrated reporter genes are used, allows for a wide choice of bacterial host strains in which the assay can be performed. These bacteria strains might present different features such as varying lipid compositions or containing non-natural lipids [64,66] allowing measurements of intramembrane TMD interactions in multiple conditions. It is also necessary to point out that, although the system requires expression of the recombinant proteins in bacteria it has served as a platform to analyze TMDs not only from prokaryotic but also from eukaryotic membrane proteins. However, in these cases the results should be analyzed carefully as the membrane environment in which the TMDs are placed might impact their interaction propensities.

5. Other wet (greasy) approaches

In this review, we have focused on those techniques currently used that could be implemented in a standard molecular biology laboratory regardless of its experience with membrane proteins and with no specialized equipment. Nonetheless, current knowledge of TMD

interactions has been achieved with the use of many other powerful techniques.

5.1. FRET, Förster Resonance Energy Transfer

Förster Resonance Energy Transfer (FRET) consists of the non-radiative (that is dipole-dipole) transfer of energy from an excited chromophore (donor) to an acceptor molecule. Although not necessary for FRET to occur, both donor and acceptor are usually fluorophores. In this case, energy transfer results in a decrease in donor emission coupled with an increase in acceptor emission. The efficiency of this energy transfer is strongly dependent on the distance between the donor and the acceptor. Therefore, it serves, indirectly, to detect PPI. The exact distance at which FRET can occur depends on the donor/acceptor used, for those used to detect PPI this distance is less than 10 nm [67]. Many molecules have been explored as FRET donors and acceptors. Nonetheless, nowadays, in the study of PPI is most common to use variants of the GFP with different ex/em wavelengths as both donors and acceptors (e.g. CFP and YFP). For a more in-depth review on the physical basis, the approaches that can be used and, recent advances in TMD-TMD interactions FRET visit [68] and [69].

FRET was first applied for the study of PPI in the late '70s and shortly after was introduced to explore membrane protein interactions [70]. Since then, FRET experiments have been fundamental in the study of TMD packing, in deciphering the effect of the hydrophobic environment on TMD-TMD association and stability, and in the development of artificial bilayers and solvents for membrane proteins. For an excellent review of the analysis of membrane protein using FRET visit [71].

5.2. SDS-PAGE

SDS electrophoresis has also proved to be a simple and qualitative approach for studying TMD-TMD interactions. Although SDS is a well-known protein denaturant, several membrane proteins maintain their oligomeric forms in the presence of SDS [59,72–75]. Although migration in SDS-PAGE does not commonly correlate with molecular weight for membrane proteins [76], it is useful for estimating oligomeric forms and for the screening of mutants. In these screenings, each residue in the interacting TMD sequence is mutated, and the extent of homo-oligomer formation in the mutants is measured by gel electrophoresis. This approach has been paradigmatically applied to GpA. The wide use of GpA as a model membrane protein is based on its intrinsic simplicity since its single TMD drives an SDS-resistant homodimerization. Thus, the dimerization process and those factors that could affect or modify it can be analyzed using SDS-PAGE. The GpA homodimer defines an interacting interface that has been extensively studied in SDS-PAGE by different techniques, including saturation mutagenesis [50,77], alanine-insertion scanning [78], and dimerization motif minimization [51,79]. Interestingly, the influence in TMD packing of hydrophobic matching between the TMD length and the hydrophobic thickness of the detergent micelles has been proved by using sodium decyl (10 carbon atoms) and tetradecyl (14 carbon atoms) sulfate as analogs of SDS (12 carbon atoms) in the polyacrylamide gels [51], extending the versatility of this straightforward technique. Despite SDS electrophoresis is not a quantitative method, a reasonable correlation has been demonstrated between the results of SDS-PAGE analyses and differences in free energies of association determined by sedimentation equilibrium analysis using analytical ultracentrifugation [80].

5.3. Analytical centrifugation

Analytical ultracentrifugation is a method used for the quantitative analysis of macromolecules in solution. It is based on the separation of proteins and protein complexes based on their size and shape under sufficiently high centrifugal force (up to 250,000 ×g) [81]. Analytical ultracentrifugation was developed for water-soluble proteins and was

quickly adapted for membrane proteins solubilized in detergent micelles [82,83]. In recent years, with the implementation of technological advances and the increased computational power and availability, an interest in analytical centrifugation has re-emerged not only for the study of soluble proteins but also for MP [84,85]. However, even with current advances, due to the nature of the technique, the study of homo- and hetero-interactions between isolated TMDs by analytical ultracentrifugation requires the fusion of the TMDs to an accompanying protein such as the staphylococcal nuclease [85–87].

6. *In silico* analysis of TMD interactions

In silico simulations of the membrane environment and the protein segments within have truly revolutionized the field of intramembrane PPI. From atomistic simulations to coarse-grained methods there has been a vast increase in the available methodology and, accordingly, in the publications [88–90]. The goal of this review is to comprehensively describe user-friendly methodologies for the study of TMD interactions and thus we are going to focus on the *in silico* techniques that have been made available by the authors through a user-friendly web interface and thus do not require specialized software or programming language.

6.1. PREDDIMER: prediction tool for an ensemble of transmembrane α-helical dimer conformations

PREDDIMER is a surface-based modeling approach for the prediction and analysis of homo- and hetero-TMD dimers and their conformational disposition [91,92]. It uses an all-atom molecular dynamics approach to identify the most probable dimer interface for two modeled TMDs implemented in an easy-to-use web server. First, using an ideal α3,6 TM helix frame and a molecular hydrophobicity potential (MHP) approach [93], the surface hydrophobicity of the sequences under study is mapped. Next, the overlap between hydrophobicity maps of two helices is systematically explored using a sliding window procedure.

The algorithm was implemented into a web server (https://preddimer.nmr.ru/preddimer/), containing two modes: *Mode 1*, for the prediction of TMD dimer structures from sequences; and *Mode 2* where the contact regions for an existing dimer structure can be analyzed. In the first case, the user should choose the modeling type (homo- *vs* hetero-dimer) and introduce the corresponding TMD sequences (with a length between 20 and 35 residues). Additionally, the server allows for the specification of the TMD orientation in the potential dimer (that is parallel or antiparallel) and the pH, which will determine the ionization state of amino acid side chains and thus influence the potential TMD-TMD interactions. Preddimer will return a series of predicted conformations. Additionally, predicted PDB structures, and 2D maps of the interaction surface are available for download. *Mode 2* requires the upload of PDB files for each interacting pair of helices. When run in *Mode 2* the algorithm will build a map for the dimer interface and prompt a results page with the same information provided in *Mode 1*.

The performance of this prediction method was tested by calculating the conformation of 11 TMD dimers whose structures were previously obtained by nuclear magnetic resonance (NMR) spectroscopy [91]. In 9 out of the 11 cases, the algorithm predicted structures similar to those experimentally obtained by NMR displaying backbone root-mean-square deviations from the references below 3 Å. The exceptions were ErbB3 and FGFR3 homo-dimers whose published structure displays anomalous packing and distorted conformations. As mentioned before the algorithm is based on an ideal α3,6 helix. Therefore, it is likely that dimers, whose interactions are only possible when deviations from canonical helices occur, are not properly predicted. Furthermore, the prediction method does not take into consideration the membrane environment. Therefore, any implications that this could have on the dimer formation are missing. Nonetheless, as the authors suggest, the predicted conformations might be used as a starting point for further molecular dynamics optimization in explicit membranes [91].

Interestingly, the different dimer configurations provided by the server might correspond to different states of TMD dimers. The first (most optimal) dimer configuration suggested by PREDDIMER for the receptor tyrosine kinase ErbB2 TMD homo-dimer corresponds to the published experimental structure, while the second one resembles an alternative structure obtained by computational methods [94].

In summary, PREDDIMER represents a user-friendly approach for the identification and analysis of TMD dimers that not only can provide structural information and guidance for experimental approaches but also can be used as a starting/reference point for more in-depth computational analysis.

6.2. TMDOCK

TMDOCK is a computational method for the modeling of parallel homodimers formed by helical TMDs (accessible *via* a web server at https://membranome.org/tmdock) [95]. TMDOCK is based on a previously published approach for estimating association free energies of α-helices in nonpolar media [96]. For the estimation of the helix-helix binding free energies three components were considered: i) the transferred energy of protein atoms from a solvent (lipid, detergent, or water) to the interior of the protein complex, ii) the hydrogen-bond, van der Waals, and electrostatic interactions within the protein interior and iii) the side-chain conformational entropy losses calculated using the discrete rotamer approximation.

Using an amino acid sequence as input, TMDOCK identifies α-helical TMDs and searches for helix association modes using the template-driven docking approach previously described. The procedure includes refinement of the dimer structures by local energy minimization and their ranking.

The method was tested with 26 experimental structures of parallel TMD homodimers of 17 single-spanning MP and 4 mutants. TMDOCK identified all high-stability native-like models, but it missed low-stability structures for three proteins. Seven proteins from the experimental dataset had alternative dimeric structures. In almost all these cases TMDOCK was able to correctly reproduce two alternative structures within the top four ranked possibilities. Importantly, TMDOCK can distinguish peptides that do not form dimers in membranes.

For optimal performance, the input sequence is expected to be longer than the TMD. The location of the TMD is automatically determined by an embedded module within TMDOCK. The output consists of a pdf file for each of the predicted models accompanied by the free energy of the helix-helix association, the ΔG_{stb} that represents the energy of the alpha-helical dimer relative to helices in water, E_{assoc} (kcal/mol) that shows the compactness of helix packing, an estimation of dimer symmetry, and inter-helical angles and distances. The potential disulfide bonds and the key residues in the interface are also highlighted.

6.3. TMHOP: Trans-Membrane Homo Oligomer Predictor

TMHOP is an automated web server recently released for the structure prediction of TM homo-oligomeric proteins. The server, based on an accurate new energy function for automated modeling and design, is available at (http://tmhop.weizmann.ac.il) free to use for academic purposes [97].

Before the design of the TMHOP algorithm, to thoroughly understand the minuses of TMD-TMD interactions, including the role of lipids and water molecules, a high-throughput experimental screen was carried out [97]. The screening, called deep sequencing ToxCAT-β-lactamase (dsTβL) was based on a variation of the previously described ToxCAT methodology [23] (Section 2) and it was used to quantify the apparent ΔG for the transfer of an amino acid from the cytosol to the *E. coli* plasma membrane [99]. Using this data, the authors developed a lipophilicity-based energy term that was integrated into the Rosetta centroid-level and all-atom functions to substitute the previous solvation model with one encoding a gradual transition from water solvation to

near and within the plasma membrane. The resulting new energy function outperforms the previous Rosetta membrane energy function [100] in three benchmarks of intramembrane PPI modeling and design: atomistic *ab initio* structure prediction starting from extended chains, prediction of mutational effects on protein stability, and sequence recovery in combinatorial sequence design [97].

TMHOP server is based on *ab initio* modeling calculations starting from fully extended chains while the majority of previous structure-prediction software use shape complementarity [101] [91,92], sequence packing motifs [102] or, comparative modeling [95], starting from canonical α-helices or experimental structures [97]. Working with *ab initio* modeling can be particularly relevant given that deviations from canonical α-helical conformation make important contributions to the packing of TMDs [97].

Importantly, it was demonstrated that TMHOP is capable of accurately predicting the structures in two-thirds of the single-span homodimers and high-order oligomers found in the Protein Data Bank (PDB). In comparison, other structure-prediction web servers like PREDDIMER [91], which algorithm operates with ideal helices (Section 6.1) [101], correctly predicted the structure for just nine of the seventeen homodimers contained in the PDB (at the time of the analysis) and was not able to predict higher-order oligomers [97].

The high accuracy models generated by TMHOP thanks to the lipophilicity preferences inferred from dsTβL energetics and the residue packing calculations in Rosetta enable accurate *ab initio* modeling of TMD oligomers, which may help to circumvent laborious and often unsuccessful experiments to determine structures of homo-oligomers [97].

6.4. THOIPA

Recently a new algorithm for the prediction of TMD homotypic interface residues has been implemented, THOIPA (TM Homodimer Interface Prediction Algorithm) [104]. THOIPA is accessible *via* a web server (thoipa.org). Alternatively, the source code can be found in github (https://github.com/bojigu/thoipapy) and a python package can be installed *via* pip (https://pypi.org/project/thoipapy/). Note that THOIPA does not predict the formation of TMD dimers, it is designed to identify which specific residues are found at the interface.

For the development of THOIPA, the authors created a database of 50 self-interacting TMDs. This database was constructed using TMD homodimers which solved structure (by NMR or X-ray crystallography) as well as information about the TMD homodimer interface obtained with reporter assays such as ToxCAT. For every residue in each of the 50 TMDs, the authors extracted properties such as conservation, polarity, co-evolution, and depth in the bilayer. Next, the features associated with homotypic TMD interactions were determined by comparing the properties of interface and non-interface residues. These position-dependent properties databases represent the core of the THOIPA machine-learning algorithm which predicts homotypic TM interface residues.

To identify the interface of a TMD homodimer the user must provide the full-length sequence of the protein of interest and the sequence of the particular TMD. The full-length sequence of the protein is used to obtain homologous sequences using BLAST. Based on the input TMD, the TMD region of each homolog is identified, extracted, and combined into multiple sequence alignments. From the multiple-sequence alignment parameters such as sequence conservation, hydrophobicity, and residues co-variation are obtained. Finally, this information is used as input for the machine learning algorithm.

The performance of TOHIPA was tested in comparison with PREDDIMER [101] and TMDOCK [95]. The authors claim that THOIPA is superior to TMDOCK and PREDDIMER predicting the most important residues in the interaction interface. However, predicting the entire interface region, THOIPA showed a modest performance.

7. Concluding remarks

The approaches described in this review provide the current means to identify and test potential TMD associations in biology. It is highly important to remember that any of these PPI assays can produce false positives due to the eventual detection of stochastic protein-protein interactions. Note that the membrane is a 2D system with a limited freedom degree and in most of the aforementioned techniques the TMDs being tested are over-expressed. Therefore, it is mandatory to include the proper controls to minimize misinterpretation of the data. Furthermore, once a TMD-TMD interaction has been found, a systematic analysis by scanning mutagenesis (ideally including different residue types) should be performed. These mutations must ideally not affect membrane insertion, sub-cellular localization, or stability of the fusion protein. If these requirements have been met, the envisioned mutations will validate the native interaction and unravel the identities of residues making up the TMD-TMD interfaces. Additionally, in any of these "in-cell" methods, it is mandatory to check that the level of expression of the different protein moieties included is comparable. Otherwise, differences in interaction-derived signals could be a mere consequence of the expression levels.

Most of the described techniques were developed with the GpA dimer in mind. Thus, these assays were designed to report TMD-TMD interactions through the reconstitution of a bipartite dimer or the activation of a homo- or heterodimer reporter. However, it is important to keep in mind that these assays cannot distinguish between dimer or higher-order oligomers.

The use of these approaches together with the emergence of new methods will significantly increase our current knowledge of the highly imbricated protein-protein interactions that take place within the cellular membranes to regulate the more relevant cellular processes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, et al., STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (D1) (2019), D607-13. Jan 8.

[3] S.A. Serapian, G. Colombo, Designing molecular spanners to throw in the protein networks, Chem Weinh Bergstr Ger. 26 (21) (2020) 4656–4670.

[4] Z. Wang, H. Sun, C. Shen, X. Hu, J. Gao, D. Li, et al., Combined strategies in structure-based virtual screening, Phys. Chem. Chem. Phys. PCCP. 22 (6) (2020) 3149–3159.

[5] S. Forli, Charting a path to success in virtual screening, Mol. Basel Switz. 20 (10) (2015) 18732–18758.

[6] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, et al., Proteomics. tissue-based map of the human proteome, Science 347 (6220) (2015), 1260419, https://doi.org/10.1126/science.1260419. PMID: 25613900.

[7] H. Kolmar, F. Hennecke, K. Götze, B. Janzer, B. Vogt, F. Mayer, et al., Membrane insertion of the bacterial signal transduction protein ToxR and requirements of transcription activation studied by modular replacement of different protein substructures, EMBO J. 14 (16) (Aug 15 1995) 3895–3904.

[8] D. Langosch, B. Brosig, H. Kolmar, H.-J. Fritz, Dimerisation of the glycophorin a transmembrane segment in membranes probed with the ToxR transcription activator, J. Mol. Biol. 263 (4) (Nov 8 1996) 525–530.

[9] W.P. Russ, D.M. Engelman, TOXCAT: a measure of transmembrane helix association in a biological membrane, Proc. Natl. Acad. Sci. 96 (3) (Feb 2 1999) 863–868.

[10] W.P. Russ, D.M. Engelman, The GxxxG motif: a framework for transmembrane helix-helix association, J. Mol. Biol. 296 (3) (Feb 25 2000) 911–919.

[11] K.R. MacKenzie, J.H. Prestegard, D.M. Engelman, A transmembrane helix dimer: structure and implications, Science 276 (5309) (Apr 4 1997) 131–133.

[12] M.G. Teese, D. Langosch, Role of GxxxG motifs in transmembrane domain interactions, Biochemistry 54 (33) (Aug 25 2015) 5125–5135.

[13] P.D. Adams, A.S. Lee, A.T. Brünger, D.M. Engelman, Models for the transmembrane region of the phospholamban pentamer: which is correct? Ann. N. Y. Acad. Sci. 16 (853) (1998) 178–185.

[14] S.M. Anderson, B.K. Mueller, E.J. Lange, A. Senes, Combination of Ca-H hydrogen bonds and van der Waals packing modulates the stability of GxxxG-mediated dimers in membranes, J. Am. Chem. Soc. 139 (44) (Nov 8 2017) 15774–15783.

[15] M.S. McClain, P. Cao, T.L. Cover, Amino-terminal hydrophobic region of helicobacter pylori vacuolating cytotoxin (VacA) mediates transmembrane protein dimerization, Infect. Immun. 69 (2) (Feb 2001) 1181–1184.

[16] J.M. Mendrola, M.B. Berger, M.C. King, M.A. Lemmon, The single transmembrane domains of ErbB receptors self-associate in cell membranes, J. Biol. Chem. 277 (7) (Feb 15 2002) 4704–4712.

[17] R. Li, R. Gorelik, V. Nanda, P.B. Law, J.D. Lear, W.F. DeGrado, et al., Dimerization of the transmembrane domain of integrin alphaIIb subunit in cell membranes, J. Biol. Chem. 279 (25) (Jun 18 2004) 26666–26673.

[18] J.P. Dawson, J.S. Weinger, D.M. Engelman, Motifs of serine and threonine can drive association of transmembrane helices, J. Mol. Biol. 316 (3) (Feb 22 2002) 799–805.

[19] J.P. Dawson, R.A. Melnyk, C.M. Deber, D.M. Engelman, Sequence context strongly modulates association of polar residues in transmembrane helices, J. Mol. Biol. 331 (1) (Aug 1 2003) 255–262.

[20] R.M. Johnson, A. Rath, C.M. Deber, The position of the gly-xxx-gly motif in transmembrane segments modulates dimer affinity, Biochem. Cell. Biol. Biochim. Biol. Cell. 84 (6) (2006) 1006–1012.

[21] R.A. Melnyk, S. Kim, A.R. Curran, D.M. Engelman, J.U. Bowie, C.M. Deber, The affinity of GXXXG motifs in transmembrane helix-helix interactions is modulated by long-range communication, J. Biol. Chem. 279 (16) (Apr 16 2004) 16591–16597.

[22] J. Zhang, T. Lazaridis, Transmembrane helix association affinity can be modulated by flanking and noninterfacial residues, Biophys. J. 96 (11) (Jun 3 2009) 4418–4427.

[23] M. Lis, K. Blumenthal, A modified, dual reporter TOXCAT system for monitoring homodimerization of transmembrane segments of proteins, Biochem. Biophys. Res. Commun. 339 (1) (Jan 6 2006) 321–324.

[24] D. Schneider, D.M. Engelman, GALLEX, a measurement of heterologous association of transmembrane helices in a biological membrane, J. Biol. Chem. 278 (5) (Jan 31 2003) 3105–3111.

[25] M. Dmitrova, G. Younès-Cauet, P. Oertel-Buchheit, D. Porte, M. Schnarr, M. Granger-Schnarr, A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in Escherichia coli, Mol. Gen. Genet. MGG 257 (2) (1998) 205–212.

[26] G. Karimova, J. Pidoux, A. Ullmann, D. Ladant, A bacterial two-hybrid system based on a reconstituted signal transduction pathway, Proc. Natl. Acad. Sci. 95 (10) (May 12 1998) 5752–5756.

[27] D. Steindorf, D. Schneider, In vivo selection of heterotypically interacting transmembrane helices: complementary helix surfaces, rather than conserved interaction motifs, drive formation of transmembrane hetero-dimers, Biochim. Biophys. Acta BBA Biomembr. 1859 (2) (2017) 245–256.

[28] B.W. Berger, D.W. Kulp, L.M. Span, J.L. DeGrado, P.C. Billings, A. Senes, et al., Consensus motif for integrin transmembrane helix association, Proc. Natl. Acad. Sci. U. S. A. 107 (2) (Jan 12 2010) 703–708.

[29] C.R. Armstrong, A. Senes, Screening for transmembrane association in divisome proteins using TOXGREEN, a high-throughput variant of the TOXCAT assay, Biochim. Biophys. Acta 1858 (11) (2016) 2573–2583.

[30] A. Bennasroune, A. Gardin, C. Auzan, E. Clauser, S. Dirrig-Grosch, M. Meira, et al., Inhibition by transmembrane peptides of chimeric insulin receptors, Cell Mol. Life Sci. 62 (18) (2005) 2124–2131.

[31] C.-D. Hu, Y. Chinenov, T.K. Kerppola, Visualization of interactions among bZIP and rel family proteins in living cells using bimolecular fluorescence complementation, Mol. Cell 9 (4) (Apr 2002) 789–798.

[32] G.S. Baird, D.A. Zacharias, R.Y. Tsien, Circular permutation and receptor insertion within green fluorescent proteins, Proc. Natl. Acad. Sci. U. S. A. 96 (20) (Sep 1999) 11241–11246.

[33] I. Ghosh, A.D. Hamilton, L. Regan, Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein, J. Am. Chem. Soc. 122 (23) (Jun 2000) 5658–5659.

[34] T. Nagai, A. Sawano, Eun Sun Park, A. Miyawaki, Circularly permuted green fluorescent proteins engineered to sense Ca2, Proc. Natl. Acad. Sci. U. S. A. 98 (6) (Mar 2001) 3197–3202.

[35] A. Tsuchisaka, A. Theologis, Heterodimeric interactions among the 1-amino-cyclopropane-1-carboxylate synthase polypeptides encoded by the arabidopsis gene family, Proc. Natl. Acad. Sci. U. S. A. 101 (8) (Feb 24 2004) 2275–2280.

[36] M.-K. Sung, W.-K. Huh, Bimolecular fluorescence complementation analysis system for in vivo detection of protein-protein interaction in Saccharomyces cerevisiae, Yeast Chichester Engl. 24 (9) (2007) 767–775.

[37] K.C. Cole, H.W. McLaughlin, D.I. Johnson, Use of bimolecular fluorescence complementation to study in vivo interactions between Cdc42p and Rdi1p of Saccharomyces cerevisiae, Eukaryot. Cell 6 (3) (Mar 2007) 378–387.

[38] A. Yano, Y. Kodama, A. Koike, T. Shinya, H.-J. Kim, M. Matsumoto, et al., Interaction between methyl CpG-binding protein and ran GTPase during cell division in tobacco cultured cells, Ann. Bot. 98 (6) (Dec 2006) 1179–1187.

[39] C.-D. Hu, A.V. Grinberg, T.K. Kerppola, Visualization of protein interactions in living cells using bimolecular fluorescence complementation (BiFC) analysis, Jan, Curr. Protoc. Cell Biol. 21 (1) (2006). Unit 21.3-21.3.21.

[40] B. Grau, M. Javanainen, M.J. García-Murria, W. Kulig, I. Vattulainen, I. Mingarro, et al., The role of hydrophobic matching on transmembrane helix packing in cells, Cell Stress. 1 (2) (Nov 2 2017) 90–106.

[41] A. Peiró, L. Martínez-Gil, S. Tamborero, V. Pallás, J.A. Sánchez-Navarro, I. Mingarro, The tobacco mosaic virus movement protein associates with but does not integrate into biological membranes, J. Virol. 88 (5) (2014) 3016–3026.

[42] L. Martínez-Gil, J.A. Sánchez-Navarro, A. Cruz, V. Pallás, J. Pérez-Gil, I. Mingarro, Plant virus cell-to-cell movement is not dependent on the transmembrane disposition of its movement protein, J. Virol. 83 (11) (2009) 5535–5543.

[43] M. Chen, C. Yan, Y. Ma, X.-E. Zhang, A tandem near-infrared fluorescence complementation system with enhanced fluorescence for imaging protein-protein interactions in vivo, Biomaterials 268 (Jan 2021), 120544.

[44] V. Andreu-Fernández, M. Sancho, A. Genovés, E. Lucendo, F. Todt, J. Lauterwasser, et al., Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes, Proc. Natl. Acad. Sci. U. S. A. 114 (2) (Jan 10 2017) 310–315.

[45] M.J. García-Murria, G. Duart, B. Grau, E. Díaz-Beneitez, D. Rodríguez, I. Mingarro, et al., Viral Bcl2s' transmembrane domain interact with host Bcl2 proteins to control cellular apoptosis, Nat. Commun. 11 (1) (2020), 27.

[46] A. Barriga, M. Morán-Lalangui, J.C. Castillo-Sánchez, I. Mingarro, J. Pérez-Gil, B. García-Álvarez, Role of pulmonary surfactant protein sp-C dimerization on membrane fragmentation: an emergent mechanism involved in lung defense and homeostasis, Biochim. Biophys. Acta Biomembr. 1863 (6) (2021 Jun 1), 183572.

[47] A. Subotic, E. Swinnen, L. Demuyser, H. De Keersmaecker, H. Mizuno, H. Tournu, et al., A bimolecular fluorescence complementation tool for identification of protein-protein interactions in Candida albicans, G3 Genes Genomes Genet. 7 (10) (2017) 3509–3520.

[48] Y.J. Shyu, H. Liu, X. Deng, C.-D. Hu, Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions, BioTechniques. 40 (1) (2006 Jan) 61–66.

[49] T. Nagai, K. Ibata, E.S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki, A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications, Nat. Biotechnol. 20 (1) (Jan 2002) 87–90.

[50] M.A. Lemmon, J.M. Flanagan, H.R. Treutlein, J. Zhang, D.M. Engelman, Sequence specificity in the dimerization of transmembrane alpha-helices, Biochemistry 31 (51) (Dec 29 1992) 12719–12725.

[51] M. Orzáez, D. Lukovic, C. Abad, E. Pérez-Payá, I. Mingarro, Influence of hydrophobic matching on association of model transmembrane fragments containing a minimised glycophorin a dimerisation motif, FEBS Lett. 579 (7) (Mar 14 2005) 1633–1638.

[52] C. Schanzenbach, F.C. Schmidt, P. Breckner, M.G. Teese, D. Langosch, Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions, Sci. Rep. 07 (7) (2017) 43476.

[53] V. Sideraki, W. Huang, T. Palzkill, H.F. Gilbert, A secondary drug resistance mutation of TEM-1 ß-lactamase that suppresses misfolding and aggregation, Proc. Natl. Acad. Sci. 98 (1) (Jan 2 2001) 283–288.

[54] J.-D. Pédelacq, S. Cabantous, T. Tran, T.C. Terwilliger, G.S. Waldo, Engineering and characterization of a superfolder green fluorescent protein, Nat. Biotechnol. 24 (1) (Jan 2006) 79–88.

[55] R. Arai, H. Ueda, A. Kitayama, N. Kamiya, T. Nagamune, Design of the linkers which effectively separate domains of a bifunctional fusion protein, Protein Eng. Des. Sel. 14 (8) (Aug 1 2001) 529–532.

[56] T.P. Hopp, K.S. Prickett, V.L. Price, R.T. Libby, C.J. March, D. Pat Cerretti, et al., A short polypeptide marker sequence useful for recombinant protein identification and purification, Bio/Technology 6 (10) (Oct 1988) 1204–1210.

[57] N.R. Cozzarelli, R.B. Kelly, A. Kornberg, A minute circular DNA from Escherichia coli 15, Proc. Natl. Acad. Sci. U. S. A. 60 (3) (Jul 1968) 992–999.

[58] J.E. Cronan, A family of arabinose-inducible Escherichia coli expression vectors having pBR322 copy control, Plasmid 55 (2) (Mar 1 2006) 152–157.

[59] M.A. Lemmon, J.M. Flanagan, J.F. Hunt, B.D. Adair, B.J. Bormann, C.E. Dempsey, et al., Glycophorin a dimerization is driven by specific interactions between transmembrane alpha-helices, J. Biol. Chem. 267 (11) (Apr 15 1992) 7683–7689.

[60] C.L. Ried, C. Scharnagl, D. Langosch, Entrapment of water at the transmembrane helix-helix Interface of quiescin sulfhydryl oxidase 2, Biochemistry 55 (9) (Mar 8 2016) 1287–1290.

[61] W. Ruan, V. Becker, U. Klingmüller, D. Langosch, The Interface between self-assembling erythropoietin receptor transmembrane segments corresponds to a membrane-spanning leucine zipper, J. Biol. Chem. 279 (5) (Jan 30 2004) 3273–3279.

[62] M. Rapp, E. Granseth, S. Seppälä, G. von Heijne, Identification and evolution of dual-topology membrane proteins, Nat. Struct. Mol. Biol. 13 (2) (Feb 2006) 112–116.

[63] C.R. Sanders, Perplexing new insight into the dynamics of the EmrE transporter, J Gen Physiol. 146 (6) (Dec 1 2015) 441–444.

[64] A. Julius, L. Laur, C. Schanzenbach, D. Langosch, BLaTM 2.0, a genetic tool revealing preferred antiparallel interaction of transmembrane helix 4 of the dual-topology protein EmrE, J. Mol. Biol. 429 (11) (Jun 2 2017) 1630–1637.

[65] A. Galarneau, M. Primeau, L.-E. Trudeau, S.W. Michnick, ß-lactamase protein fragment complementation assays as in vivo and in vitro sensors of protein–protein interactions, Nat. Biotechnol. 20 (6) (Jun 2002) 619–622.

[66] I. Shibuya, Metabolic regulation and biological functions of phospholipids in Escherichia coli, Prog. Lipid Res. 31 (3) (Jan 1 1992) 245–299.

[67] T. Förster, Zwischenmolekulare energiewanderung und fluoreszenz, Ann. Phys. 437 (1–2) (1948) 55–75.

[68] D.W. Piston, G.-J. Kremers, Fluorescent protein FRET: the good, the bad and the ugly, Trends Biochem. Sci. 32 (9) (Sep 2007) 407–414.

[69] K. Okamoto, Y. Sako, Recent advances in FRET for the study of protein interactions and dynamics, Curr. Opin. Struct. Biol. 46 (2017) 16–23.

[70] S.J. Morris, T.C. Südhof, D.H. Haynes, Calcium-promoted resonance energy transfer between fluorescently labeled proteins during aggregation of chromaffin granule membranes, Biochim. Biophys. Acta 693 (2) (Dec 22 1982) 425–436.

[71] A.S. Khadria, A. Senes, Fluorophores, environments and quantification techniques in the analysis of transmembrane helix interaction using FRET, Biopolymers 104 (4) (Jul 2015) 247–264.

[72] J. Fujii, K. Maruyama, M. Tada, D.H. MacLennan, Expression and site-specific mutagenesis of phospholamban. studies of residues involved in phosphorylation and pentamer formation, J. Biol. Chem. 264 (22) (Aug 5 1989) 12950–12955.

[73] H. Splitt, D. Meuser, I. Borovok, M. Betzler, H. Schrempf, Pore mutations affecting tetrameric assembly and functioning of the potassium channel KcsA from streptomyces lividans, FEBS Lett. 472 (1) (Apr 21 2000) 83–87.

[74] H. Gratkowski, J.D. Lear, W.F. DeGrado, Polar side chains drive the association of model transmembrane peptides, Proc. Natl. Acad. Sci. U. S. A. 98 (3) (Jan 30 2001) 880–885.

[75] R. Laage, J. Rohde, B. Brosig, D. Langosch, A conserved membrane-spanning amino acid motif drives homomeric and supports heteromeric assembly of presynaptic SNARE proteins, J. Biol. Chem. 275 (23) (Jun 9 2000) 17481–17487.

[76] A. Rath, M. Glibowicka, V.G. Nadeau, G. Chen, C.M. Deber, Detergent binding explains anomalous SDS-PAGE migration of membrane proteins, Proc. Natl. Acad. Sci. U. S. A. 106 (6) (Feb 10 2009) 1760–1765.

[77] M. Orzáez, J. Salgado, A. Giménez-Giner, E. Pérez-Payá, I. Mingarro, Influence of proline residues in transmembrane helix packing, J. Mol. Biol. 335 (2) (Jan 9 2004) 631–640.

[78] I. Mingarro, P. Whitley, G.V. Heijne, M.A. Lemmon, Ala-insertion scanning mutagenesis of the glycophorin a transmembrane helix: a rapid way to map helix-helix interactions in integral membrane proteins, Protein Sci. 5 (7) (Jul 1 1996) 1339–1341.

[79] M. Orzáez, E. Pérez-Payá, I. Mingarro, Influence of the C-terminus of the glycophorin a transmembrane fragment on the dimerization process, Protein Sci. Publ. Protein Soc. 9 (6) (Jun 2000) 1246–1253.

[80] K.G. Fleming, D.M. Engelman, Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants, Proc. Natl. Acad. Sci. U. S. A. 98 (25) (Dec 4 2001) 14340–14344.

[81] T.R. Patel, D.J. Winzor, D.J. Scott, Analytical ultracentrifugation: a versatile tool for the characterisation of macromolecular complexes in solution, Methods San Diego Calif. 15 (95) (2016 Feb) 55–61.

[82] C. Tanford, J.A. Reynolds, Characterization of membrane proteins in detergent solutions, Biochim. Biophys. Acta BBA - Rev. Biomembr. 457 (2) (1976 Oct 26) 133–170.

[83] C. Tanford, Y. Nozaki, J.A. Reynolds, S. Makino, Molecular characterization of proteins in detergent solutions, Biochemistry 13 (11) (May 21 1974) 2369–2376.

[84] E.C. Smith, S.E. Smith, J.R. Carter, S.R. Webb, K.M. Gibson, L.M. Hellman, et al., Trimeric transmembrane domain interactions in paramyxovirus fusion proteins roles in protein folding, stability and function 288 (50) (2013) 35726–35735.

[85] S.R. Webb, S.E. Smith, M.G. Fried, R.E. Dutch, Transmembrane domains of highly pathogenic viral fusion proteins exhibit trimeric association in vitro, mSphere 3 (2) (2018) 25.

[86] A.Z. Ebie, K.G. Fleming, Dimerization of the erythropoietin receptor transmembrane domain in micelles, J. Mol. Biol. 366 (2) (Feb 16 2007) 517–524.

[87] A.M. Stanley, K.G. Fleming, The transmembrane domains of ErbB receptors do not dimerize strongly in micelles, J. Mol. Biol. 347 (4) (Apr 8 2005) 759–772.

[88] L.M. Simpson, B. Taddese, I.D. Wall, C.A. Reynolds, Bioinformatics and molecular modelling approaches to GPCR oligomerization, Curr. Opin. Pharmacol. 10 (1) (Feb 1 2010) 30–37.

[89] Al-Balas Q.A. Al-Shar'i NA, Molecular dynamics simulations of adenosine receptors: advances, applications and trends, Curr. Pharm. Des. 25 (7) (2019) 783–816.

[90] T. Mori, N. Miyashita, W. Im, M. Feig, Y. Sugita, Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms, Biochim. Biophys. Acta Biomembr. 1858 (7 Part B) (2016) 1635–1651. Jul 1.

[91] A.A. Polyansky, A.O. Chugunov, P.E. Volynsky, N.A. Krylov, D.E. Nolde, R. G. Efremov, PREDDIMER: a web server for prediction of transmembrane helical dimers, Bioinformatics 30 (6) (Mar 15 2014) 889–890.

[92] A.A. Polyansky, P.E. Volynsky, R.G. Efremov, Multistate organization of transmembrane helical protein dimers governed by the host membrane, J. Am. Chem. Soc. 134 (35) (Sep 5 2021) 14390–14400.

[93] R.G. Efremov, A.O. Chugunov, T.V. Pyrkov, J.P. Priestle, A.S. Arseniev, E. Jacoby, Molecular lipophilicity in protein modeling and drug design, Curr. Med. Chem. 14 (4) (2007) 393–415.

[94] S.J. Fleishman, J. Schlessinger, N. Ben-Tal, A putative molecular-activation switch in the transmembrane domain of erbB2, Proc. Natl. Acad. Sci. 99 (25) (Dec 10 2002) 15937–15940.

[95] A.L. Lomize, I.D. Pogozheva, TMDOCK: an energy-based method for modeling a-helical dimers in membranes, J. Mol. Biol. 429 (3) (Feb 3 2017) 390–398.

[96] A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, Quantification of helix-helix binding affinities in micelles and lipid bilayers, Protein Sci. Publ. Protein Soc. 13 (10) (2004 Oct) 2600–2612.

[97] J.Y. Weinstein, A. Elazar, S.J. Fleishman, A lipophilicity-based energy function for membrane-protein modelling and design, Aug 28 [cited 2020 Apr 9], PLoS Comput. Biol. 15 (8) (2019), https://www.ncbi.nlm.nih.gov/pmc/articles/PM C6736313/.

[99] A. Elazar, J. Weinstein, I. Biran, Y. Fridman, E. Bibi, S.J. Fleishman, Mutational scanning reveals the determinants of protein insertion and association energetics in the plasma membrane. Shan Y, editor, elife 5 (2016 Jan 29), e12125.

[100] V. Yarov-Yarovoy, J. Schonbrun, D. Baker, Multipass membrane protein structure prediction using Rosetta, Proteins 62 (4) (2006) 1010–1025. Mar 1.

[101] A.A. Polyansky, A.O. Chugunov, P.E. Volynsky, N.A. Krylov, D.E. Nolde, R. G. Efremov, PREDDIMER: a web server for prediction of transmembrane helical dimers, Bioinformatics 30 (6) (2014) 889–890. Mar 15.

[102] H. Cao, M.C.K. Ng, S.A. Jusoh, H.K. Tai, S.W.I. Siu, TMDIM: an improved algorithm for the structure prediction of transmembrane domains of bitopic dimers, J. Comput. Aided Mol. Des. 31 (9) (2017) 855–865. Sep 1.

[104] Y. Xiao, B. Zeng, N. Berner, D. Frishman, D. Langosch, Teese M. George, Experimental determination and data-driven prediction of homotypic transmembrane domain interfaces, Comput. Struct. Biotechnol. J. 18 (2020) 3230–3242.

[105] D.S. Goodsell, L. Autin, A.J. Olson, Illustrate: software for biomolecular illustration, Structure 27 (11) (2019 Nov 5) 1716–1720, e1.

8.2. Annex II

Duart, G., Mingarro, I., Martinez-Gil, L., 2021c. *The importance of transmembrane domain interactions in the viral control of apoptosis.* ***Mol Cell Oncol*** 8, 1911290. **https://doi.org/10.1080/23723556.2021.1911290**

The importance of transmembrane domain interactions in the viral control of apoptosis

Gerard Duart, Ismael Mingarro & Luis Martinez-Gil

Published online: 03 May 2021.

Submit your article to this journal ↗

Article views: 100

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

AUTHOR'S VIEWS

Check for updates

The importance of transmembrane domain interactions in the viral control of apoptosis

Gerard Duart [iD], Ismael Mingarro [iD], and Luis Martinez-Gil [iD]

Department of Biochemistry and Molecular Biology, Institut de Biotecnologia i Biomedicina, Universitat de València, Burjassot, Spain

ABSTRACT
Viral control of apoptosis occurs through the expression of viral encoded anti-apoptotic B-cell lymphoma 2 (BCL2) analogs. These proteins are thought to restrain apoptosis by interacting with cellular BCL2 family members. We identified that protein-protein interactions between cellular and viral BCL2 transmembrane domains are crucial for the viral protein's function.

In multi-cellular organisms, tissue development and homeostasis relies on tight control of programmed cell death. Furthermore, miss-regulation of apoptosis, or any other cellular mechanism that participates in the control of cell fate, has a strong impact on developed organisms as it usually leads to cancer, auto-immunity, or neurodegeneration among other disorders. Accordingly, programmed cell death is heavily regulated. This control relies primarily on the B-cell lymphoma 2 (BCL2) protein family. The BCL2 family, consisting of approximately 20 proteins, includes pro-survival, pro-apoptotic, and apoptosis activators. Pro- and anti–apoptotic BCL2 family members share four sequence homology domains (BCL2 Homology domain 1-4, BH1–4). On the other hand, activators of the BH3-only subgroup solely have the BH3 domain. Besides, most BCL2 proteins have a transmembrane domain (TMD) in the carboxyl-terminal (Ct) end that facilitates insertion into the target membrane.

Interactions among BCL2 proteins are crucial for the regulation of apoptosis. Anti-apoptotic BCL2 proteins inhibit the activation of pro-apoptotic members of this family through direct interaction or sequestering BH3-only activators. After an apoptotic stimulus, pro-apoptotic proteins and/or BH3-only activators will be released and, in turn, induce cell death through mitochondrial membrane permeabilization. These interactions among BCL2 family members were thought to occur through soluble domains, particularly through the BH3 domain. However, a seminal work proposed that interactions between pro- and anti-apoptotic TMDs can occur.[1]

Apoptosis and autophagy also play a major role in the control and clearance of infectious diseases. It is generally accepted that apoptosis-induction is, in most infections, beneficial to the host. Eliminating infected cells mitigates the propagation of the infection and stimulates an appropriate immune response.[2] To elude the cell death mediated antiviral response viruses have developed multiple strategies to control apoptosis. These include masking of internal cellular sensors,

caspase regulation, signaling cascade modulation, and mimicking of BCL2 regulators with viral homologs known as viral BCL2s (vBCL2s; singular, vBCL2).[3,4]

There is a structural homology between vBCL2s and their cellular counterparts. As most cellular BCL2 (cBCL2) many vBCL2s present a hydrophobic amino acid stretch on their Ct end. Using an *in vitro* assay based on the *E. coli* leader peptidase, an assay that allows for quantitative description of the membrane insertion capability of short sequences,[5] we found that the Ct hydrophobic regions of herpesviral (HHV4, HHV8, and BOHV4) and poxviral (VACV, MYXV, and ORFV) vBCL2s can insert into endoplasmatic reticulum (ER)-derived membranes as TMD despite a theoretical low hydrophobicity score for some of these regions.[6] To avoid confusion, here we use the viral acronym to refer to the vBCL2 protein.

Next, we assessed whether the TMD of these vBCL2s exhibits any sort of self-association properties. For this purpose we used two approaches, a bimolecular fluorescent complementation (BiFC) assay adapted for the study of intramembrane interactions[7] and BLaTM, a genetic tool designed to study TMD–TMD interactions in the bacterial membrane.[8] Our results indicated that all vBCL2 TMDs can form homo-oligomers in mitochondrial membranes. Additionally, we investigated the potential TMD–TMD interactions between vBCL2 and pro-, anti-apoptotic, and BH3-only cBCL2s.

Our BiFC-based screening revealed that most viral TMDs can interact with multiple cellular TMDs. However, the particularities of these intramembrane protein-protein interaction networks varied from virus to virus, revealing distinctive mechanisms of action. Of note, we observed similar connection circuits among closely related viruses. These similarities could not have been inferred by the analysis of the TMDs sequences, which suggests a structural pattern underlying the sequence that governs intramembrane interactions.
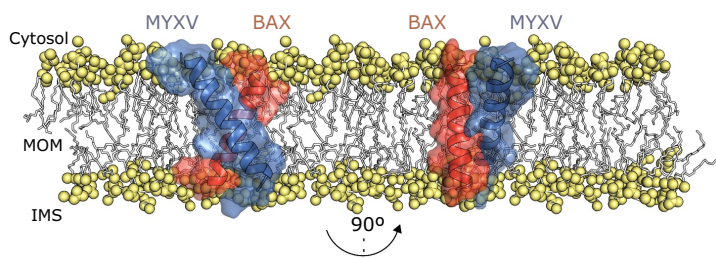
Figure 1. Interaction between BAX and myxovirus vBCL2 transmembrane domains. Model of a putative dimer between MYXV transmembrane domain (TMD, blue) and Bax TMD (red) obtained with the PredDIMER algorithm.[9] The position of the mitochondrial outer membrane (MOM) was adjusted manually. The cytosolic and the inner membrane side (IMS) of the MOM are indicated. TMDs are shown with a cartoon (for the peptide backbones) and a surface (for side chains) representations.

An in-depth analysis of the TMD-TMD interactions between MYXV and the pro-apoptotic protein Bax on one hand, and HHV8 and BCL2 on the other, revealed that these intramembrane interactions are governed by ridge–groove arrangements created by an adequate disposition of large and small residues, where glycine residues play a key role by maximizing intimate contacts (Figure 1).

Next, to analyze whether the observed TMD–TMD interactions are required to control cellular apoptosis, we transfected HeLa cells with vBCL2 either with or without the TMD. We completed our study by including chimeras in which the TMD of each vBCL2 protein was replaced by the TMD of TOMM20, a mitochondrial protein that cannot establish TMD-TMD interactions with any cBCL2. Additionally, cells were either treated with doxorubicin or infected with VACV to induce apoptosis. Our results demonstrated that, once the TMD was removed the vBCL2 proteins could not promote survival or stop apoptosis (measured by Trypan blue staining and flow cytometry using propidium iodide staining and phosphatidylserine labeling (FITC-Annexin V), or by Caspase 3/7 levels). Similarly, the chimeras carrying the TMD of TOMM20 could not control apoptosis. These results suggest that TMD–TMD hetero-oligomerizations are crucial for modulating cell death regardless of the nature of the apoptotic stimulus.

Our work expands our knowledge about how viruses interact with their host and point to the membrane hydrophobic core as a new playground for host-viral interactions. Furthermore, these results increase our understanding of how viruses control cellular apoptosis and how apoptosis is regulated in the cell. The necessity of TMD-TMD interactions for successful apoptosis inhibition opens a new avenue for the development of therapeutic drugs against viral pathogens characterized by short- and long-term deregulation of programmed cell death.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

ORCID

Gerard Duart http://orcid.org/0000-0002-9012-3164
Ismael Mingarro http://orcid.org/0000-0002-1910-1229
Luis Martinez-Gil http://orcid.org/0000-0002-9076-7760

References

1. Andreu-Fernández V, Sancho M, Genovés A, Lucendo E, Todt F, Lauterwasser J, Funk K, Jahreis G, Pérez-Payá E, Mingarro I, et al. Bax transmembrane domain interacts with prosurvival Bcl-2 proteins in biological membranes. Proc Natl Acad Sci U S A. 2017;114:310–315. doi:10.1073/pnas.1612322114.
2. Häcker G. Apoptosis in infection. Microbes Infect. 2018;20:552–559. doi:10.1016/j.micinf.2017.10.006.
3. Galluzzi L, Brenner C, Morselli E, Touat Z, Kroemer G. Viral control of mitochondrial apoptosis. PLoS Pathog. 2008;4: e1000018. doi:10.1371/journal.ppat.1000018.
4. Polčic P, Mentel M, Gavurníková G, Bhatia-Kiššová I. To keep the host alive - the role of viral Bcl-2 proteins. Acta Virol. 2017;61:240–251. doi:10.4149/av_2017_302.
5. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, Von Heijne G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature. 2005;433:377–381. doi:10.1038/nature03216.
6. García-Murria MJ, Duart G, Grau B, Diaz-Beneitez E, Rodríguez D, Mingarro I, Martínez-Gil L. Viral BCL2s' transmembrane domain interact with host BCL2 proteins to control cellular apoptosis. Nat Commun. 2020;11:6056. doi:10.1038/s41467-020-19881-9.
7. Grau B, Javanainen M, García-Murria MJ, Kulig W, Vattulainen I, Mingarro I, Martínez-Gil L. The role of hydrophobic matching on transmembrane helix packing in cells. Cell Stress. 2017;1:90–106. doi:10.15698/cst2017.11.111.
8. Schanzenbach C, Schmidt FC, Breckner P, Teese MG, Langosch D. Identifying ionic interactions within a membrane using BLaTM, a genetic tool to measure homo- and heterotypic transmembrane helix-helix interactions. Sci Rep. 2017;7:43476. doi:10.1038/srep43476.
9. Polyansky AA, Chugunov AO, Volynsky PE, Krylov NA, Nolde DE, Efremov RG. PREDDIMER: a web server for prediction of transmembrane helical dimers. Bioinformatics. 2014;30:889–890. doi:10.1093/bioinformatics/btt645.

8.3. Annex III

Papers not included in this thesis

García-Murria, M.J., Expósito-Domínguez, N., **Duart, G.**, Mingarro, I., Martinez-Gil, L., 2019. *A Bimolecular Multicellular Complementation System for the Detection of Syncytium Formation: A New Methodology for the Identification of Nipah Virus Entry Inhibitors.* **Viruses** 11, 229. https://doi.org/10.3390/v11030229

Lucendo, E., Sancho, M., Lolicato, F., Javanainen, M., Kulig, W., Leiva, D., **Duart, G.**, Andreu-Fernández, V., Mingarro, I., Orzáez, M., 2020. *Mcl-1 and Bok transmembrane domains: Unexpected players in the modulation of apoptosis.* **Proc Natl Acad Sci USA**. https://doi.org/10.1073/pnas.2008885117