Membrane-protein topology

Gunnar von Heijne

Abstract | In the world of membrane proteins, topology defines an important halfway house between the amino-acid sequence and the fully folded three-dimensional structure. Although the concept of membrane-protein topology dates back at least 30 years, recent advances in the field of translocon-mediated membrane-protein assembly, proteome-wide studies of membrane-protein topology and an exponentially growing number of high-resolution membrane-protein structures have given us a deeper understanding of how topology is determined and of how it evolves.

Topology
A specification of the number of transmembrane helices and their in and/or out orientations across the membrane in a membrane protein.

Fold space
The abstract space of all protein folds.

'Knobs-into-holes' geometry
The classic mode of helix–helix packing in which side-chains on one helix fit into spaces between side chains on the opposite helix.

Center for Biomembrane Research and Stockholm Bioinformatics Center, Department of Biochemistry and Biophysics, Stockholm University, SE-10691 Stockholm.
e-mail: gunnar@dbb.su.se
doi:10.1038/nrm2063

In the year 2000, it all seemed so simple: membrane proteins of the helix-bundle type were built from long, hydrophobic α-helices that were orientated more or less perpendicularly to the membrane plane[1]. The topology of a membrane protein was just a description, listing the segments of the polypeptide chain that form the transmembrane helices and their orientation relative to the membrane.

But now, 6 years on and 60 high-resolution membrane-protein structures later, the picture is not so simple. Membrane-embedded helices can be short, long, kinked or interrupted in the middle of the membrane, they can cross the membrane at oblique angles, lie flat on the surface of the membrane, or even span only a part of the membrane and then turn back, forming so-called re-entrant loops. Therefore, depending on one's view, one can adopt a more or less complicated definition of topology. Even the concept of a transmembrane helix is not entirely clear-cut — exactly how far must a helix reach towards the water-interface regions on both sides of the membrane to qualify as transmembrane?

As our picture of the underpinnings of membrane-protein structure has become more complex, we have also come to appreciate that the topology of membrane proteins can evolve in interesting ways, and these might not even be the same for all molecules that have identical amino-acid sequences. Individual transmembrane helices in a protein might vary in their capability to insert efficiently into the membrane, giving rise to a statistical distribution of topologies with different numbers of transmembrane helices. Dual-topology proteins insert into the membrane in two opposite orientations with an approximate 1:1 stoichiometry. Also, homologous proteins can evolve to insert with opposite orientations, or homologous proteins can fuse to form structures with two antiparallel membrane-spanning domains. It has even been shown that transmembrane helices can dynamically reorient across the membrane in response to drastic changes in lipid composition. However, whether such reorientation phenomena are also part of the normal functioning of membrane proteins is less clear.

In this review, I summarize the recent developments in our understanding of membrane-protein topology and structure. I aim to provide a (reasonably) integrated view of membrane-protein biogenesis, which starts from the basic process of helix integration into a membrane and moves through membrane-protein topology to the full three-dimensional (3D) structure. These issues will be presented in an evolutionary context in an attempt to identify the main types of molecular-scale evolutionary event that have helped shape the extent of membrane proteomes.

What the structures say

At the time of writing, the Protein Data Bank held more than 100 high-resolution structures of integral membrane proteins of the helix-bundle type. The number of known membrane-protein structures is growing exponentially and doubles every ~3 years, trailing the statistics for water-soluble proteins by ~15 years[2].

Yet, despite this encouraging trend, we are still only scratching the surface of the fold space of membrane proteins, and each new structure brings new surprises. As the crystallographers' attention has shifted from the rock-solid electron- and proton-conducting membrane-protein complexes that are involved in photosynthesis and respiration to the more inherently flexible channel and transport proteins, we have begun to appreciate the large dynamic changes in helix–helix packing interactions, and even in local helical–non-helical structural transitions that can occur in membrane proteins. It can no longer be taken for granted that all transmembrane helices are straight, orientated approximately perpendicularly to the membrane, or that they pack with each other according to simple 'knobs-into-holes' geometries.

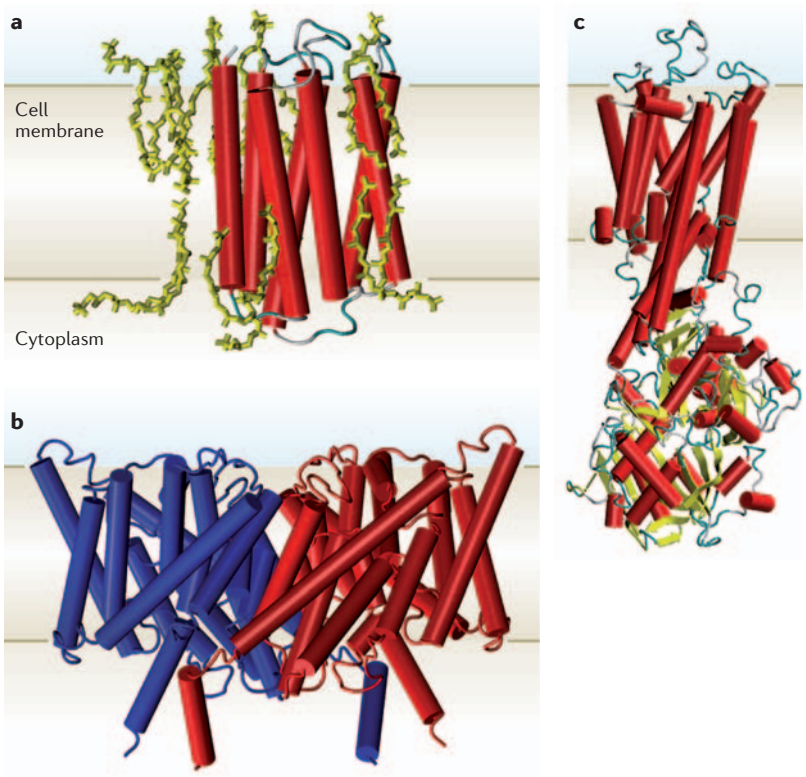

Figure 1 | **Helix-bundle membrane proteins. a** | Bacteriorhodopsin[83] (Protein Data Bank (PDB) accession code 2BRD). The seven transmembrane helices are shown in red and co-crystallizing membrane phospholipids are shown in yellow. **b** | The *Escherichia coli* ClC Cl^-/H^+ antiporter[6] (PDB accession code 1KPK) is a homodimer (one subunit is shown in blue and the other in red). There is one Cl^- channel in each monomer. **c** | Bovine Ca^{2+}-ATPase[10] (PDB accession code 1T5S) coloured according to secondary structure (helices are red; β-strands are yellow). For movies that display rotating versions of the bacteriorhodopsin and ClC structures as well as the mechanistic cycle of Ca^{2+}-ATPase, see Supplementary information S1–S3 (movies).

FIG. 1 (see also Supplementary information S1–S3 (movies)) shows a comparison of three remarkably different structures: bacteriorhodopsin, the bacterial chloride channel ClC and bovine Ca^{2+}-ATPase. Bacteriorhodopsin (FIG. 1a) is a classic seven-helix bundle with a covalently bound retinal in the middle. It functions as a light-driven proton pump: small, light-induced movements in its transmembrane helices entice protons to translocate across the membrane against an electrochemical gradient[3]. The helices lie almost straight in the membrane and pack with typical knobs-into-holes packing angles[4,5].

The chloride channel ClC (FIG. 1b) is a homodimer, and each monomer has a Cl^--specific channel[6,7]. As well as containing several standard transmembrane helices, the structure contains both long, steeply inclined helices and short, closely spaced pairs of helices that penetrate only halfway through the membrane (re-entrant loops). There are also stretches of non-helical structure deep within the membrane that are largely buried between the transmembrane helices.

The Ca^{2+}-ATPase (FIG. 1c and Supplementary information S3 (movie)) switches between dramatically different conformations during its ATP-driven pumping cycle[8–11]. Extramembranous ATP-binding and phosphorylation domains move relative to each other in response to activities at the ATP-binding site and cation binding in the membrane, and the relative rotation of the so-called actuator domain alone exceeds 110°. Coupled to these changes, individual transmembrane helices move by >8 Å perpendicularly to the membrane, and entire turns of helices can fold and unfold. Small-molecule transporters such as lactose permease and the glycerol-3-phosphate transporter are also thought to undergo extensive repacking of their transmembrane helices during substrate transport[12,13]. Likewise, the S4 transmembrane helix in the voltage-sensor domain of voltage-dependent ion channels has been proposed to move by as much as 15–20 Å across the membrane in response to changes in the membrane potential[14].

Despite this continually growing list of structural variability and dynamics, one early generalization that has held up well is the distinct distribution profiles of the different amino acids along the transmembrane helices (FIG. 2). The frequency of hydrophobic residues (Ala, Ile, Val and Leu) peaks in the middle of the membrane, that of the aromatic residues Tyr and Trp (but not Phe) peaks in the lipid–water interface regions, and charged and polar residues are largely absent from the membrane interior[15]. Furthermore, positively charged residues have a skewed distribution and are more frequently found in cytoplasmic, compared to non-cytoplasmic, parts of the protein — the positive-inside rule[16,17].

In summary, membrane proteins can be dynamic entities, with their transmembrane helices changing position in the membrane, partially folding and unfolding, and repacking during a reaction cycle. However, there is no clear case in which the basic topology of a protein changes as the protein cycles between different conformational states. So in what follows, for the most part, I will use the classic concept of membrane-protein topology; that is, helix-bundle membrane proteins can be described as assemblies of transmembrane helices that are connected by loops that reach at least into the lipid-headgroup region of a membrane.

Membrane-protein assembly

The topology of a membrane protein is, in most cases, determined during the initial insertion of the polypeptide chain into the membrane. With few exceptions, the insertion is mediated by a translocon — a molecular gate-keeper that allows nascent polypeptide chains to pass across or integrate into lipid membranes. The best understood translocons are the Sec61 translocon in the rough endoplasmic reticulum (ER) and its bacterial homologue, the SecYEG translocon.

The Sec61 translocon forms a protein-conducting channel across the ER membrane and is used by all secretory proteins to gain entry into the secretory pathway. It is also used by membrane proteins, but with an interesting twist in that their transmembrane helices are not translocated all the way across the membrane, but rather are shunted sideways into the lipid bilayer, presumably through a 'lateral gate' in the wall of the translocon[18,19] (FIG. 3a). A basic characteristic of all transmembrane

Retinal
The light-sensitive cofactor in bacteriorhodopsin that absorbs photons and triggers a conformational change in the protein.

Electrochemical gradient
The combined pH and electrostatic gradient across a membrane.

S4 transmembrane helix
A positively charged transmembrane helix that forms part of the voltage-sensor domain in voltage-dependent ion channels.

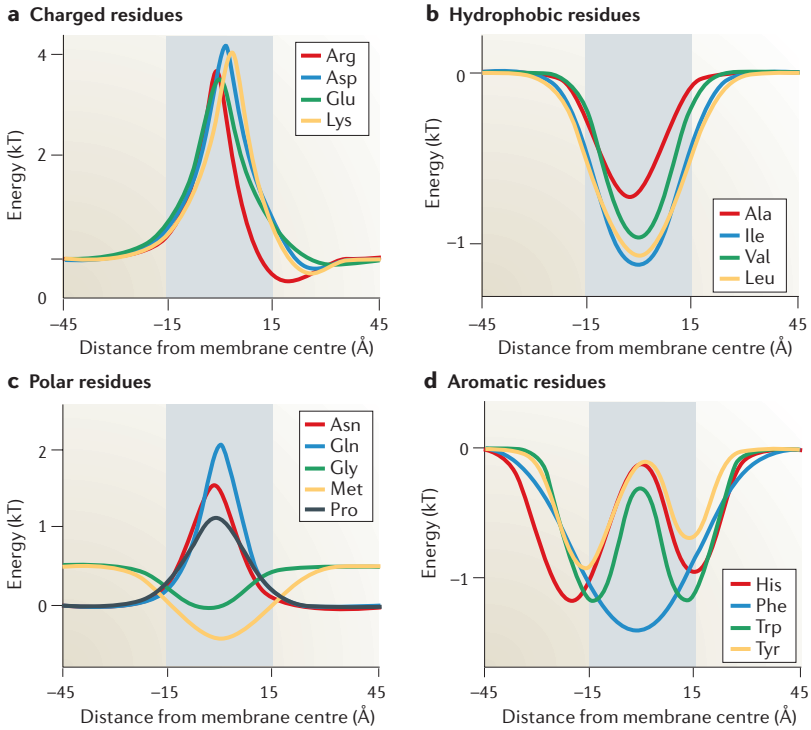

Figure 2 | **Different amino acids have distinct preferences for different parts of the membrane.** The graphs show statistical free energies of membrane insertion, which were calculated from a set of high-resolution X-ray structures[15], for: (**a**) charged residues; (**b**) hydrophobic residues; (**c**) polar residues; and (**d**) aromatic residues. The higher the energy at a particular position, the lower the probability that the amino acid is present at that position. The grey shading highlights the position of the apolar core of the membrane. Modified with permission from REF. 15 © (2005) John Wiley & Sons Ltd.

helices is therefore their capability to engage with the lateral gate and thereby ensure their proper integration into the membrane.

The high-resolution X-ray structure of a purified archaeal Sec61 translocon[20] and lower-resolution electron-microscopy structures of ribosome-bound mammalian and bacterial translocons[21,22] have provided tantalizing, but as-yet incomplete, pictures of the translocon in action. The X-ray structure of the translocon shows a channel with a narrow central constriction that is closed from the periplasmic end by a 'plug helix' (FIG. 3b). The overall structure is clam-shell like, and it has been proposed that a lateral gate can open towards the lipid bilayer at the mouth of the clam-shell[20]. Transmembrane helices in a translocating nascent polypeptide are thought to exit the translocon through this gate when they insert into the membrane. How much the channel opens up in its active, ribosome-bound state is unclear, and therefore it is also not possible to say how many transmembrane helices can be sequestered in the channel at the same time. The stoichiometry of the active, ribosome-bound translocon is hotly debated, and different models for the structure of the active Sec61 complex have been proposed[20,21,22]. In one model, the active translocon is proposed to be a single Sec61 complex with a nascent polypeptide chain translocating through a narrow channel, whereas in another it is proposed to be a front-to-front Sec61 dimer with two lateral gates partly open and facing each other and with a much wider central channel.

Chemical crosslinking of ribosome-bound nascent chains to the Sec61 translocon has been extensively used to characterize the immediate environment that is encountered by a transiting polypeptide[18]. By placing photo-activatable probes at defined positions in different transmembrane helices, the insertion process can be followed in considerable detail. For aquaporin-4, a protein with 6 transmembrane helices, it was recently shown that the helices contact the Sec61α subunit of the ER translocon in a strict N-to-C-terminal succession: first helix-1, then helix-2, and so on[23]. Furthermore, any given helix was only found to leave the translocon and insert into the lipid bilayer once the next helix had entered the translocon. Similar data have been reported for the transmembrane helices-1–3 of bovine opsin[24]. It is not yet known how general such an ordered N-to-C-terminal-insertion pathway is, but it fits nicely with the idea that transmembrane helices exit the translocon through the lateral gate one by one, or possibly in pairs. The roles played by proteins that are associated with the Sec61 translocon, such as translocation-associated membrane protein (TRAM) in the ER and YidC in *Escherichia coli*, are still unclear, but it seems likely that they can have chaperoning functions, perhaps as temporary storage sites for transmembrane helices that need to be assembled with partner helices before they can insert efficiently into the lipid membrane[19,25].

Another view on how the Sec61 translocon handles transmembrane segments has been provided by studies in which the translocon is challenged by specially designed polypeptide segments that are variably hydrophobic and that are embedded in a larger model protein that serves as cargo for the translocon. Such studies have been carried out in my laboratory and have shown a surprisingly close correlation between the statistical distribution of different amino acids in the transmembrane helices of known X-ray structures[15], the partitioning free energy of amino acids in polar and non-polar solvents[26], and the capability of the different amino acids to promote or reduce the insertion of model hydrophobic segments into the ER membrane[27,28]. This indicates a model in which the lateral gate in the translocon allows the transiting nascent chain to sample the surrounding lipid bilayer, and subsequent membrane insertion is driven by the thermodynamic partitioning of hydrophobic polypeptide segments into the lipid[27,29]. If this model can be substantiated, it will provide the sought-after physical underpinnings for the hydrophobicity-based topology-prediction algorithms that are currently in use.

The mechanisms of action of other translocons, such as the translocator inner membrane (TIM) complexes that guide proteins either across or into the inner membrane of mitochondria[30], are less well understood. Yet, the basic structural characteristics of membrane proteins that use these translocons do not differ in any obvious ways from those of the proteins that are handled by the Sec61 or SecYEG translocons. It is possible that translocon-mediated recognition of the transmembrane helices is always based on similar principles.

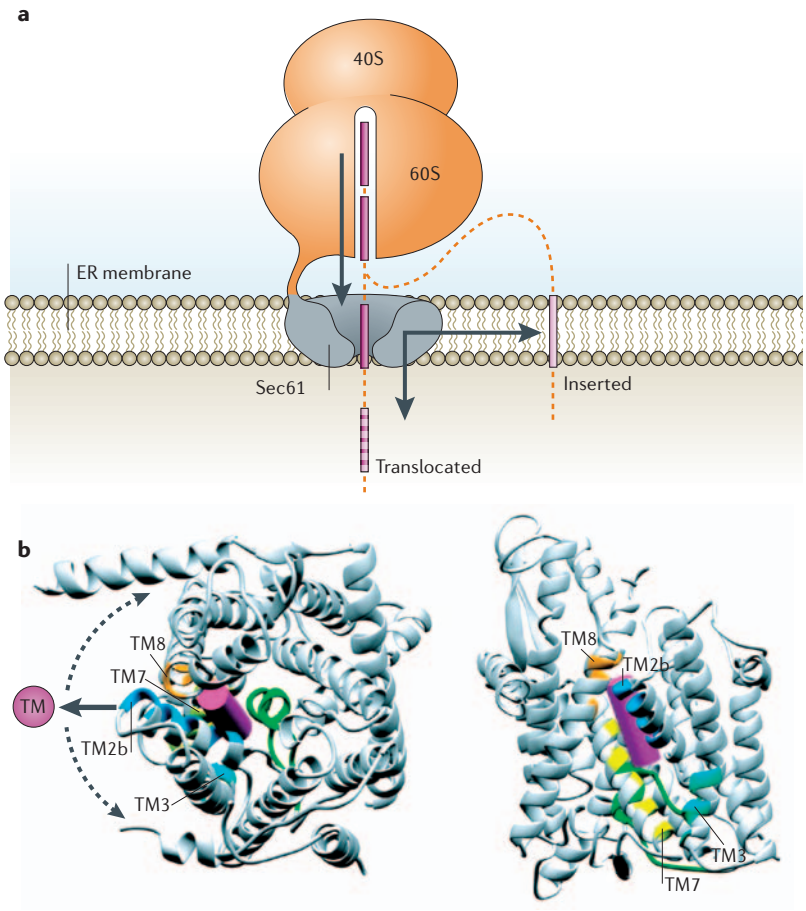

Figure 3 | **Recognition of a transmembrane helix by the Sec61 translocon.** **a** | A non-polar amino-acid segment (purple) in a co-translationally translocating nascent polypeptide chain (dashed orange line) is shown moving progressively down through the tunnel in the 60S ribosomal subunit, in step with chain elongation. Depending on the hydrophobicity of the segment, the Sec61 translocon can either shunt this segment laterally into the surrounding membrane, which results in the formation of a transmembrane helix (inserted state), or allow it to pass through the membrane into the lumen of the endoplasmic reticulum (ER) (translocated state). **b** | The Sec61 translocon from *Methanococcus jannaschii* as viewed from above (left) and in the plane of the membrane (right). The purple cylinder depicts the hypothetical position of a nascent-chain transmembrane helix that is about to move into the membrane through the 'lateral gate' between helices TM2b–3 (both shown in light blue) and TM7–8 (shown in yellow and orange). The channel is closed from the periplasmic side by the green 'plug helix'; this plug is thought to move out of the way when the ribosome binds to the translocon. Additional structural changes in the translocon upon ribosome binding are likely, but none have yet been clearly defined. Panel **b** is reproduced with permission from REF. 20 © (2004) Macmillan Magazines Ltd.

Membrane proteomes

Ever since reasonably reliable topology-prediction schemes came on the market, attempts have been made to identify membrane proteomes and to computationally predict membrane-protein topologies. Current estimates indicate that 20–30% of all the predicted open reading frames in a typical genome encode membrane proteins with one or more transmembrane helices[31]. Of these proteins, most have been predicted to have both their N and C termini in the cytoplasm, and this has recently been confirmed experimentally for both *E. coli* and *Saccharomyces cerevisiae* membrane proteomes[32,33]. This might reflect a preferred mechanism of membrane insertion, with independently inserting helical hairpins as a basic element of membrane-protein structure.

Not surprisingly, there is a strong correlation between topology and function (FIG. 4). Energy-driven small-molecule transport proteins with ~6 or ~12 transmembrane helices dominate both *E. coli* and *S. cerevisiae* membrane proteomes. Two-component signalling receptors (chemoreceptors) with 2 transmembrane helices are numerous in *E. coli*, and in both *E. coli* and *S. cerevisiae* there are a large number of proteins with 2–7 helices that have no annotated function. In animals (but not in plants), the G-protein-coupled-receptor superfamily — the members of which have seven transmembrane helices and an N_{out}–C_{in} orientation (that is with an extracytoplasmic N-terminus and a cytoplasmic C terminus) — has undergone a massive expansion. In mammals, this superfamily alone accounts for nearly 5% of all protein-coding genes[34].

Current topology-prediction methods are all based on the classic view of membrane-protein topology that was discussed in the introductory section, and they happily ignore complicating issues such as re-entrant loops, short breaks in a helix and helices that lie flat on the membrane surface. Some progress towards more complete prediction algorithms has recently been reported[35–37], but much work remains to be done before these more advanced types of prediction reach the level of accuracy of the simple topology-prediction schemes.

Opposite and dual topologies

In the large-scale *E. coli* and *S. cerevisiae* topology-mapping studies (see above), topology-reporter proteins were fused to the C termini of ~600 membrane proteins from each organism, which made it possible to infer the location of each protein's C terminus (cytoplasmic or extracytoplasmic)[32,33]. A subsequent search for homologous proteins in these data sets revealed that most homologous membrane proteins have the same C-terminal location. However, homologous proteins with opposite C-terminal orientations were found in a few interesting cases, which raises the question of how such proteins might have evolved.

There are two possible ways that homologous proteins can have oppositely orientated C termini: there can be an extra C-terminal transmembrane helix in one protein but not in the other, or the two proteins can be oppositely orientated in the membrane. Examples of both cases can be found in the *E. coli* and *S. cerevisiae* data sets. For example, in *E. coli*, the putative Arg and ornithine antiporter YdgI has an extra C-terminal transmembrane helix compared to its closest homologues, PotE and YjdE. On the other hand, YdgQ and YdgL, both with six transmembrane helices, are oppositely orientated, as are the four-helix proteins YdgE and YdgF. In *S. cerevisiae*, the four-helix protein Ygl263w is oppositely orientated compared to the eight other members of the same protein family. In all cases, these oppositely orientated homologous proteins follow the positive-inside rule — that is, each protein is orientated such that the side containing the higher number of positively charged Arg and Lys residues faces the cytoplasm.

Helical hairpin
A pair of closely spaced transmembrane helices that is connected by a short extracytoplasmic loop.

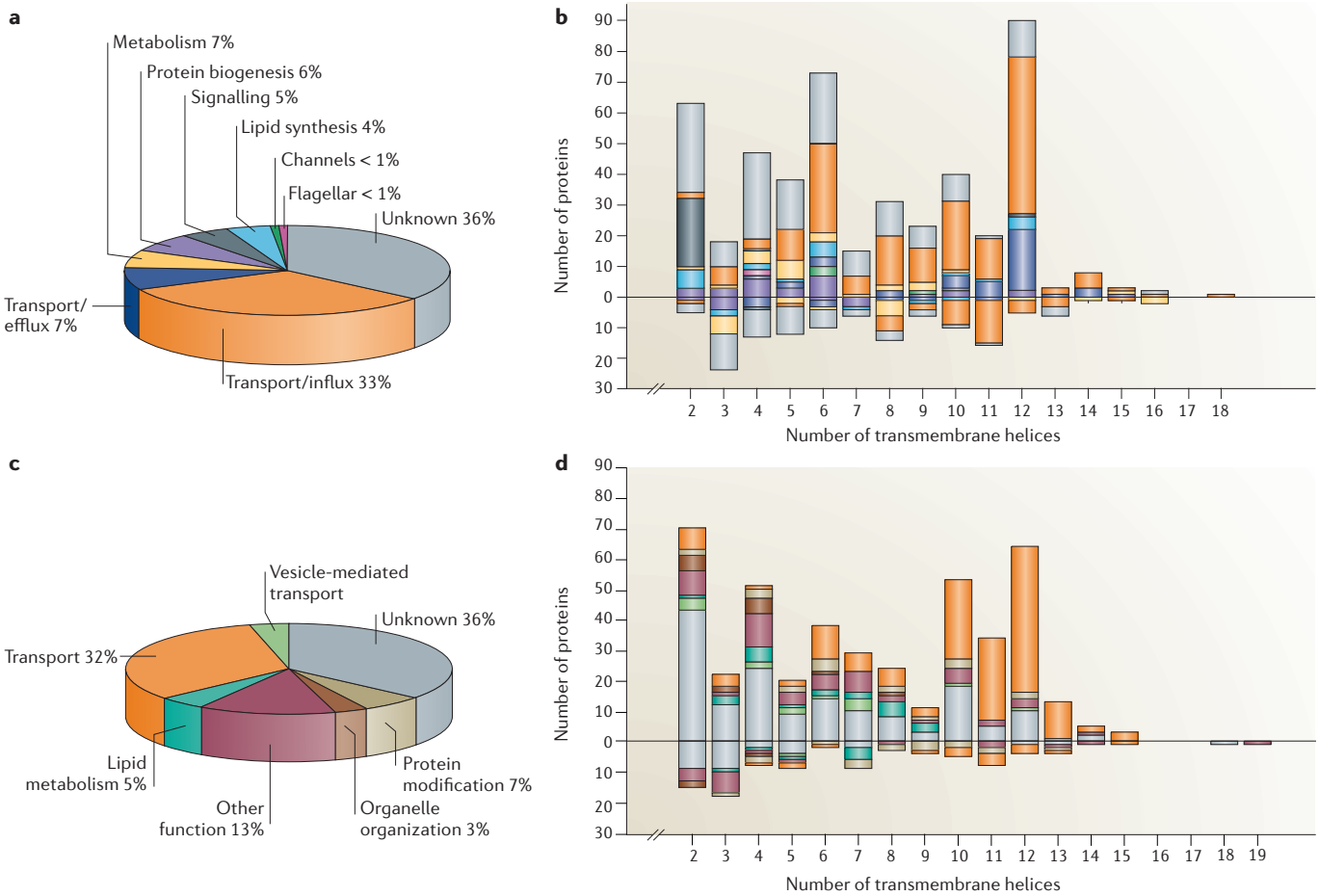

Figure 4 | **The distribution of topologies in membrane proteomes. a** | The proportion of membrane-spanning proteins in the proteome of *Escherichia coli* associated with different cellular functions. **b** | The number of membrane proteins in each functional group is shown expressed against the number of transmembrane helices. Proteins with a cytoplasmic C terminus (C_{in}) are plotted upwards and those with an extracytoplasmic C terminus (C_{out}) are plotted downwards. Functional groups of proteins are colour coded according to the upper schemes. Proteins with a single predicted transmembrane segment were not included in these studies and they are therefore not shown in these plots. **c** | Represents the same information as panel **a**, but for the *Saccharomyces cerevisiae* proteome. **d** | Represents the same information as panel **b**, but for the *S. cerevisiae* proteome. Panels **a** and **b** are modified with permission from REF. 32 © (2005) American Association for the Advancement of Science. Panels **c** and **d** are modified with permission from REF. 33 © (2005) National Academy of Sciences, USA.

The most remarkable topology variation that has been uncovered so far is dual topology. Dual topology refers to proteins that are 'undecided' in terms of their overall orientation in the membrane, and can insert in 2 opposite orientations with an approximate 1:1 stoichiometry. Although the existence of dual-topology proteins is still controversial[38], recent experimental and comparative genomics data provide strong support for at least five cases of dual-topology proteins in *E. coli*[39]. These are camphor-resistance protein (CrcB), the small multidrug resistance (SMR) proteins EmrE and SugE, and YdgC and YnfA, two proteins of unknown function. All five are small proteins with four transmembrane helices, and each has only a small difference in the number of positively charged residues between its two sides. As would be expected for dual-topology proteins, all five proteins can be pushed towards either the N_{in}–C_{in} or the N_{out}–C_{out} orientation by the introduction or the removal of single positively charged residues in the loops that connect the transmembrane helices. A recent high-resolution X-ray structure of EmrE shows it as a homodimer composed of two oppositely orientated molecules[40] (FIG. 5; see also Supplementary information S4 (movie)). This is again consistent with a dual-topology structure, although it is still unclear how well this X-ray structure fits with a lower-resolution structure that was determined by electron crystallography[41,42].

Another example of a dual-topology protein is ductin, a protein with four transmembrane helices that is found both in V-type ATPases and as a component of the connexin channel in gap junctions. These two forms of the protein have opposite membrane orientations, although it is unclear to what extent the distribution of positively charged residues influences the topology[43].

V-type ATPase
Ion-pumping ATP synthase located in intracellular organelles.

Connexin channel
A component of gap junctions.

Gap junction
A structure that connects neighbouring cells and allows ions and small molecules to pass between cells.

Figure 5 | **The candidate dual-topology protein EmrE from *Escherichia coli*.** The candidate dual-topology protein EmrE from *Escherichia coli*[40] (Protein Data Bank accession code 2F2M). The two identical, but oppositely orientated, chains are shown in blue and red. The essential and strictly conserved Glu14 residues are shown in yellow. For a movie that displays a rotating version of this structure, see Supplementary information S4 (movie).

Multiple topologies

When tested individually, not all transmembrane helices in multispanning membrane proteins insert efficiently into a membrane[44]. If, as proposed above, membrane insertion is fundamentally a thermodynamic partitioning process, then the existence of such non-inserting helices in natural proteins implies that helices can already interact with each other during the membrane-insertion step, possibly forming helical hairpins that partition into the membrane *en bloc*[45]. Another possibility is that a weakly inserting helix retains this characteristic in its natural context, such that it gives rise to a protein with multiple topologies.

An intriguing case of multiple topologies is that of the scrapie prion protein (PrP), which has both an inefficient N-terminal signal peptide and an inefficiently inserted transmembrane helix[46–48]. This combination can give rise to at least four different topologies — a cytoplasmic form, a fully secreted form and two oppositely orientated membrane-spanning forms — that might have different disease-causing potencies[49,50].

Another example of multiple topologies is provided by the multidrug transporter P-glycoprotein[44]. It has been reported that different segments of this protein can span the membrane, which results in 2 different topological forms, both with 12 transmembrane helices.

In these cases, topological diversity is caused by one or more marginally hydrophobic, and therefore inefficiently inserted, transmembrane segments. Another way to induce multiple topological forms of a protein is by manipulations based on the positive-inside rule. 'Frustrated' membrane proteins have been designed in which the more highly positively charged loops can only remain in the cytoplasm if one of the hydrophobic segments is forced to not insert across the membrane[51]. Conversely, it has been shown that a weakly hydrophobic segment can be forced to become a transmembrane segment by placing it between two transmembrane helices that have the same orientational preferences[52]. Other factors that have been shown to affect topology include the length of an N-terminal hydrophobic signal-anchor segment[53] (longer segments favour an N_{out}–C_{in} orientation), rapid folding of an N-terminal globular domain[54] and *N*-linked glycosylation of loops that become transiently exposed to the ER lumen during membrane-protein assembly[53,55].

Dynamic topologies

Throughout the preceding discussion, a tacit assumption has been that the topology, once formed, is stable. In other words, transmembrane helices do not flip in and out of the membrane, and they do not invert their orientation relative to the lipid bilayer. Considering the slow rate of spontaneous lipid flip–flop in pure lipid bilayers, it would seem even less probable that transmembrane helices could undertake similar gymnastics. However, there are a few reports that this might in fact occur.

The most dramatic scenario would be proteins that flip between different topologies as part of a reaction cycle. SecG, a component of the bacterial SecYEG translocon, has been proposed to do just this during preprotein translocation[56,57], although this has recently been disputed[58].

On a less extreme note, several membrane proteins seem to reorientate one or more of their transmembrane helices post-translationally to reach their final (stable) topology. In aquaporin-1, a protein with six transmembrane helices and two re-entrant loops, the second and fourth helices do not adopt a transmembrane orientation during the co-translational membrane-integration stage. Instead, the third helix reorientates post-translationally, bringing the second and fourth helix into their proper transmembrane positions[59] (FIG. 6a). There are also clear indications that there is a crucial time window within which an N-terminal transmembrane helix can reorientate co-translationally while still in the translocon[53].

Viral proteins also provide examples of unusual topological behaviour. A rather striking case is that of the large envelope glycoprotein of the hepatitis B virus, which is initially inserted into the ER membrane with its N-terminal so-called pre-S domain located in the cytoplasm. During a protracted, chaperone-dependent post-translational maturation process, the pre-S domain is then translocated across the membrane in about 50% of the molecules, resulting in a mixed topology with either three or four transmembrane helices for this protein[60–62] (FIG. 6b). A similar post-translational translocation of the positively charged N-terminal tail in the phage λ S107 holin protein, driven by a decrease in the membrane potential, has been suggested to lie behind the 'lysis clock' function that is encoded by this phage[63].

The hepatitis C virus envelope glycoprotein provides an example of another reorientation phenomenon that has also been observed in other viral proteins[64]. This viral envelope glycoprotein is a polyprotein that is composed of the proteins E1, E2 and p7. E1 has a cleavable N-terminal signal peptide and a hydrophobic C-terminal membrane anchor that also functions as a signal peptide for E2, and

Scrapie prion protein
An aggregation-prone protein that causes the Scrapie disease in sheep and goats.

Signal peptide
An N-terminal extension on secretory proteins that serves to target a protein to the Sec61 translocon.

Lipid flip–flop
The process whereby a lipid molecule flips between the two leaflets of a lipid bilayer.

Polyprotein
Proteins made as a single polypeptide chain that is cleaved into smaller proteins by cellular proteases.

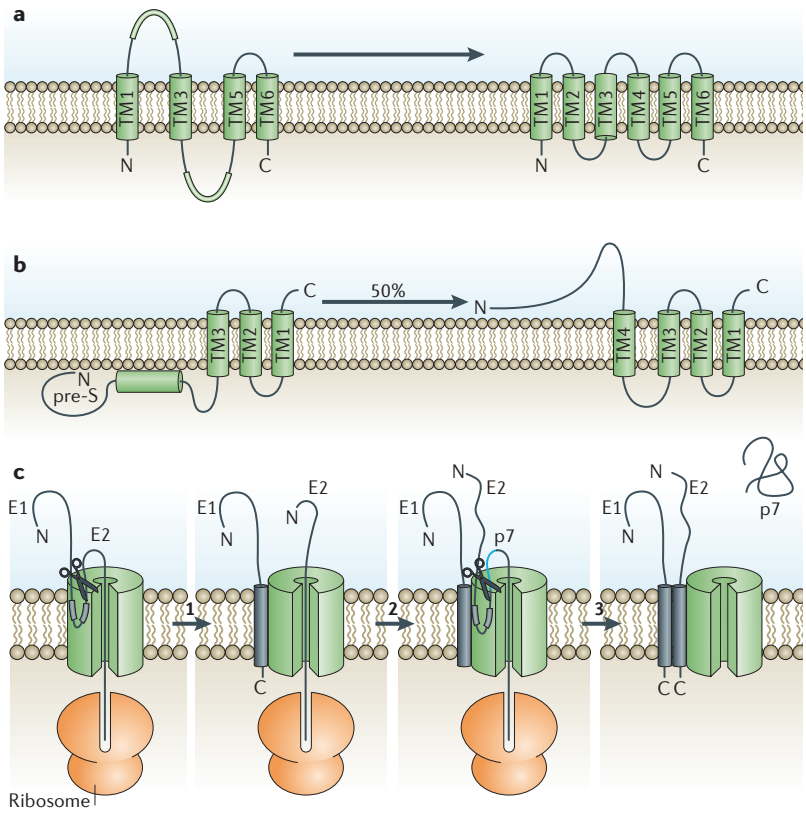

Figure 6 | **Dynamic topologies. a** | Aquaporin-1. The second and fourth transmembrane helices insert properly across the membrane only on reorientation of the third transmembrane helix[59]. **b** | The hepatitis B virus large envelope glycoprotein. The N-terminal so-called pre-S domain translocates across the endoplasmic reticulum (ER) membrane in a slow, post-translational process in ~50% of the molecules[62]. **c** | The hepatitis C virus E1–E2–p7 polyprotein attached to a ribosome and associated with a translocon. The C-terminal hydrophobic segment in E1 initially inserts as a hairpin with a short cytoplasmic turn, in this way serving as a signal peptide for the E2 protein. After cleavage by the lumenal signal peptidase (scissors), the newly generated C terminus of E1 flips back across the membrane, converting the hairpin into a single transmembrane helix (step **1**). In the same way, the hydrophobic segment at the end of E2 serves as a signal peptide for p7 (step **2**). Since p7 lacks a hydrophobic segment, after its cleavage, it is fully translocated across the membrane (step **3**)[64]. Panel **b** is modified with permission from REF. 62 © (2003) National Academy of Sciences, USA. Panel **c** is modified with permission from REF. 64 © (2002) Macmillan Magazines Ltd.

E2 has a C-terminal anchor that doubles up as a signal peptide for the p7 protein. The E1 and E2 C-terminal anchor and signal-peptide segments apparently initially insert as short 'inverted' helical hairpins with tight cytoplasmic loops that separate the signal peptide from the preceding hydrophobic stretch. Following the cleavage of the signal peptide by the lumenally located signal peptidase, the C-terminal end of the cleaved signal peptide flips back across the membrane, resulting in the formation of a single transmembrane helix at the C-terminus of both E1 and E2 (FIG. 6c).

Last, work on the *E. coli* lactose transporter, a protein with 12 transmembrane helices, has shown that this protein's topology is dramatically affected when it is expressed in cells genetically engineered to lack the lipid phosphatidylethanolamine in their inner membrane. Even more remarkably, this protein reverts to its normal topology when phosphatidylethanolamine or the foreign lipid monoglucosyldiacylglycerol is reintroduced[65–69]. Whether such lipid-induced reorientations are also possible under normal physiological conditions is not known. At the very least, these observations indicate that the interactions between charged residues in a protein and the surface charge of the membrane can in part explain the positive-inside rule[69,70], even if other factors such as the membrane potential and specific residues in the Sec61 translocon can also contribute[71,72].

There is not yet a clear case of a membrane protein that has dynamic transmembrane-helix flip–flop as part of a functional cycle. However, the post-translational reorientation of one or more helices as part of a folding or maturation process has been shown for a number of proteins.

Topology evolution

Compared to water-soluble proteins, topology provides an extra dimension that membrane proteins can evolve. Topology can evolve, for example, by the addition or removal of terminal or internal transmembrane helices, by gene fusion or fission and by the wholesale inversion of membrane orientation. So far, only a few studies have addressed topology evolution *per se*, but some interesting trends have already emerged.

The main mechanism for topology evolution is internal gene duplication[73], giving rise to internally 2-fold symmetric 3D structures[6,12,20,74,75]. Duplications can be either complete, which leads to a doubling of the number of transmembrane helices, or partial, and extra transmembrane helices can be added to the N terminus or the C terminus of a protein. A particularly interesting kind of internal duplication occurs when a protein with an odd number of transmembrane helices is fully duplicated, as this creates a protein in which the two homologous domains cannot both retain their original overall orientation in the membrane and at the same time cannot insert all their transmembrane helices across the membrane (see later).

Although membrane proteins frequently evolve by internal gene duplication, they rarely evolve by domain recombination involving non-homologous membrane domains[76]. Instead, membrane proteins often engage in noncovalent interactions with other membrane proteins to form multidomain protein complexes. However, domain recombination between a membrane-integral domain and one or more water-soluble domains is common[76], and the identity of the soluble domain can sometimes aid in the prediction of the topology of the membrane domain[77].

Topology inversion — that is, the evolution of homologous proteins with the same number of transmembrane helices but with the opposite orientation in the membrane — has been documented in a few cases. The most common way to achieve this seems to be the redistribution of positively charged residues from one side of the protein to the other. This can be achieved by protein engineering[78,79], and now has also been discovered in natural proteins both in bacteria[39] and *S. cerevisiae*[33].

Domain recombination
An evolutionary process in which pre-existing protein domains are fused in new combinations, creating multidomain, multifunctional polypeptides.

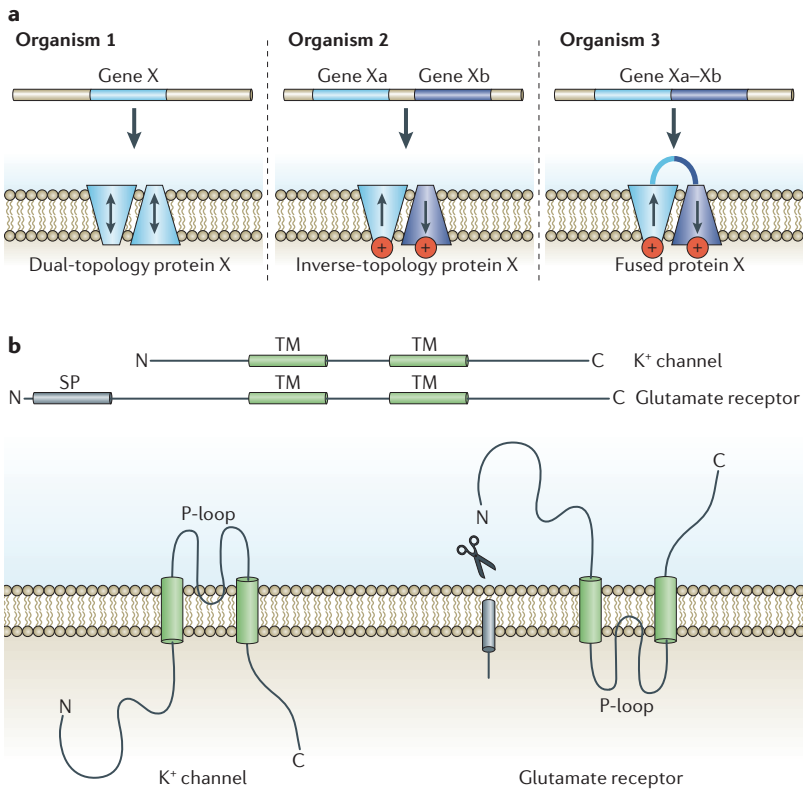

Figure 7 | **Topology evolution. a** | A possible scenario for the evolution of a dual-topology protein first into two oppositely orientated proteins, encoded by two homologous genes, and then into an internally duplicated protein with two oppositely orientated membrane domains, as a result of a fusion event between homologous genes[39]. **b** | The glutamate receptor. Glutamate receptors have an N-terminal signal peptide (SP) that K^+ channels do not have[82]. The signal peptide induces the translocation of the large N-terminal ligand-binding domain across the endoplasmic reticulum membrane as it is translated by the ribosome. The two transmembrane domains (TM) that form the ion channel therefore insert in opposite orientations compared with K^+ channels. The pore-lining P-loop faces the cytoplasm in glutamate receptors but faces the outside of the cell in K^+ channels. The signal peptide is removed from the glutamate receptor by a lumenal signal peptidase. Panel **a** is modified with permission from REF. 84

An illustrative example is provided by the SMR-protein family, which comprises the *E. coli* candidate dual-topology protein EmrE (see above) as well as the *E. coli* proteins YdgE and YdgF, and a large number of SMR proteins in other bacteria. EmrE is active as a homodimer[80], whereas YdgE and YdgF need to be co-expressed to confer drug resistance[81]. EmrE, YdgE and YdgF all have four predicted transmembrane helices, but whereas EmrE seems to have dual topology, YdgE and YdgF are oppositely orientated in the membrane[39], in agreement with the distribution of positively charged residues in the two proteins. Moreover, EmrE is encoded by an isolated gene on the *E. coli* chromosome, whereas YdgE and YdgF are encoded by a pair of neighbouring genes. A possible scenario is therefore that the YdgE–YdgF pair resulted from a gene duplication, in which an ancestral, EmrE-like, dual-topology protein evolved into two oppositely orientated homologues (FIG. 7a). Sequence searches have even uncovered a family of proteins (Pfam accession number DUF606) with five predicted transmembrane helices. This family includes: candidate dual-topology proteins, which are encoded by 'singleton' genes; oppositely orientated homologues that are encoded by neighbouring genes; and internally duplicated ten-transmembrane-helix proteins in which the first and second halves are predicted to have opposite orientations[39].

A second way to reorientate a protein is to add a cleavable signal peptide to the N terminus of a membrane protein that originally has a cytoplasmic N terminus. This can cause translocation of the N-terminal part of the protein to the extracytoplasmic side with the concomitant inversion of the orientation of the transmembrane region. A striking example of this is a prokaryotic glutamate receptor — a K^+ channel that has its N-terminal ligand-binding domain located on the outside of the cell rather than in the cytoplasm, and in which the channel domain has the opposite membrane orientation compared to most other K^+ channels[82] (FIG. 7b).

The addition of an N-terminal signal peptide would seem to be a simple evolutionary mechanism for topology inversion, but, in most cases, the positive-inside rule would probably prevent the inversion of the transmembrane helices, leading to 'topological frustration'[51]. Topologically frustrated membrane proteins have conflicting topological information in different parts of the polypeptide, and the compromise solution is either to leave one or more hydrophobic segments out of the membrane[51] or to force a non-hydrophobic segment to insert across the membrane[52]. Topology inversion by the addition (or removal) of an N-terminal signal peptide is therefore probably restricted to proteins with only a small number of transmembrane helices, and even for these it can require subsequent sequence optimization according to the positive-inside rule to achieve a unique orientation.

Conclusions

Even if some of the recently determined membrane-protein structures have to some extent 'muddied the waters', it is remarkable how far two simple concepts — the hydrophobic character of transmembrane helices and the positive-inside rule — go in providing an explanatory framework for membrane-protein topology. However, it is also slightly discouraging (or encouraging, depending on your view of life) to realize how little we know about the molecular details of how transmembrane helices are recognized and of how they are orientated in the correct way across the membrane by the different translocons in the cell. Another unattained goal, for which progress promises to be rapid, is to reach a quantitative understanding of the energetics of membrane-protein insertion and folding *in vivo*. It would, at the least, be intellectually satisfying if our topology-prediction methods and helix-packing algorithms could be based on direct measurements of interaction energies rather than on statistically-derived parameters, even if their performances weren't dramatically improved. After all, life is all physical chemistry, right?

1. von Heijne, G. Recent advances in the understanding of membrane protein assembly and structure. *Quart. Rev. Biophys.* **32**, 285–307 (2000).
2. White, S. H. The progress of membrane protein structure determination. *Protein Sci.* **13**, 1948–1949 (2004).
3. Edmonds, B. W. & Luecke, H. Atomic resolution structures and the mechanism of ion pumping in bacteriorhodopsin. *Front. Biosci.* **9**, 1556–1566 (2004).
4. Bowie, J. U. Helix packing in membrane proteins. *J. Mol. Biol.* **272**, 780–789 (1997).
5. Langosch, D. & Heringa, J. Interaction of transmembrane helices by a knobs-into-holes geometry reminiscent of soluble coiled coils. *Proteins* **31**, 150–159 (1998).
6. Dutzler, R., Campbell, E. B., Cadene, M., Chait, B. T. & MacKinnon, R. X-ray structure of a ClC chloride channel at 3. 0 Å reveals the molecular basis of anion selectivity. *Nature* **415**, 287–294 (2002).
7. Dutzler, R., Campbell, E. B. & MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **300**, 108–112 (2003).
8. Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–611 (2002).
9. Olesen, C., Sorensen, T. L., Nielsen, R. C., Moller, J. V. & Nissen, P. Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science* **306**, 2251–2255 (2004).
10. Sorensen, T. L., Moller, J. V. & Nissen, P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* **304**, 1672–1675 (2004).
11. Toyoshima, C., Nomura, H. & Tsuda, T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* **432**, 361–368 (2004).
12. Abramson, J. *et al.* Structure and mechanism of the lactose permease of *Escherichia coli*. *Science* **301**, 610–615 (2003).
13. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. *Science* **301**, 616–620 (2003).
14. Ruta, V., Chen, J. & MacKinnon, R. Calibrated measurement of gating-charge arginine displacement in the KvAP voltage-dependent K^+ channel. *Cell* **123**, 463–475 (2005).
15. Ulmschneider, M. B., Sansom, M. S. & Di Nola, A. Properties of integral membrane protein structures: derivation of an implicit membrane potential. *Proteins* **59**, 252–265 (2005).
16. von Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **5**, 3021–3027 (1986).
17. Nilsson, J., Persson, B. & von Heijne, G. Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* **60**, 606–616 (2005).
18. Alder, N. N. & Johnson, A. E. Cotranslational membrane protein biogenesis at the endoplasmic reticulum. *J. Biol. Chem.* **279**, 22787–22790 (2004).
19. Rapoport, T. A., Goder, V., Heinrich, S. U. & Matlack, K. E. Membrane-protein integration and the role of the translocation channel. *Trends Cell. Biol.* **14**, 568–575 (2004).
20. van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44 (2004).
The first, and so far only, high-resolution structure of a translocon.
21. Morgan, D. G., Menetret, J. F., Neuhof, A., Rapoport, T. A. & Akey, C. W. Structure of the mammalian ribosome-channel complex at 17 Å resolution. *J. Mol. Biol.* **324**, 871–886 (2002).
22. Mitra, K. *et al.* Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438**, 318–324 (2005).
A single-particle electron-microscopy structure of a ribosome–translocon complex.
23. Sadlish, H., Pitonzo, D., Johnson, A. E. & Skach, W. R. Sequential triage of transmembrane segments by Sec61α during biogenesis of a native multispanning membrane protein. *Nature Struct. Mol. Biol.* **12**, 870–878 (2005).
A detailed analysis of the order in which different transmembrane helices exit the Sec61 translocon.
24. Ismail, N., Crawshaw, S. G. & High, S. Active and passive displacement of transmembrane domains both occur during opsin biogenesis at the Sec61 translocon. *J. Cell Sci.* **119**, 2826–2836 (2006).
25. Luirink, J., von Heijne, G., Houben, E. & de Gier, J. W. Biogenesis of inner membrane proteins in *Escherichia coli*. *Annu. Rev. Microbiol.* **59**, 329–355 (2005).
26. White, S. H. & Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struc.* **28**, 319–365 (1999).
27. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377–381 (2005).
The first quantitative analysis of the translocon-mediated recognition of transmembrane helices.
28. Hessa, T., White, S. H. & von Heijne, G. Membrane insertion of a potassium channel voltage sensor. *Science* **307**, 1427 (2005).
29. Heinrich, S., Mothes, W., Brunner, J. & Rapoport, T. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102**, 233–244 (2000).
30. Rehling, P., Brandner, K. & Pfanner, N. Mitochondrial import and the twin-pore translocase. *Nature Rev. Mol. Cell. Biol.* **5**, 519–530 (2004).
31. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model. Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
32. Daley, D. O. *et al.* Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* **308**, 1321–1323 (2005).
The first experimental proteome-wide study of membrane-protein topology.
33. Kim, H., Österberg, M., Melén, K. & von Heijne, G. A global topology map of the *Saccharomyces cerevisiae* membrane proteome. *Proc. Natl Acad. Sci. USA* **103**, 11142–11147 (2006).
34. Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharm.* **63**, 1256–1272 (2003).
35. Granseth, E., Viklund, H. & Elofsson, A. ZPRED: predicting the distance to the membrane center for residues in α-helical membrane proteins. *Bioinformatics* **22**, e191–e196 (2006).
36. Viklund, H., Granseth, E. & Elofsson, A. Reentrant regions in α-helical transmembrane proteins are divided in 3 structural classes and abundant in small residues. *J. Mol. Biol.* **361**, 591–603 (2006).
37. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **101**, 959–963 (2004).
38. Ninio, S., Elbaz, Y. & Schuldiner, S. The membrane topology of EmrE — a small multidrug transporter from *Escherichia coli*. *FEBS Lett.* **562**, 193–196 (2004).
39. Rapp, M., Seppälä, S., Granseth, E. & von Heijne, G. Identification and evolution of dual topology membrane proteins. *Nature Struct. Mol. Biol.* **13**, 112–116 (2006).
40. Pornillos, O., Chen, Y. J., Chen, A. P. & Chang, G. X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* **310**, 1950–1953 (2005).
41. Ubarretxena-Belandia, I., Baldwin, J. M., Schuldiner, S. & Tate, C. G. Three-dimensional structure of the bacterial multidrug transporter EmrE shows it is an asymmetric homodimer. *EMBO J.* **22**, 6175–6181 (2003).
42. Tate, C. G. Comparison of three structures of the multidrug transporter EmrE. *Curr. Opin. Struct. Biol.* **16**, 457–464 (2006).
43. Dunlop, J., Jones, P. C. & Finbow, M. E. Membrane insertion and assembly of ductin: a polytopic channel with dual orientations. *EMBO J.* **14**, 3609–3616 (1995).
44. Sadlish, H. & Skach, W. R. Biogenesis of CFTR and other polytopic membrane proteins: new roles for the ribosome-translocon complex. *J. Membr. Biol.* **202**, 115–126 (2004).
45. Meindl-Beinker, N. M., Lundin, C., Nilsson, I., White, S. H. & von Heijne, G. Asn- and Asp-mediated interactions between transmembrane helices during translocon-mediated membrane protein assembly. *EMBO Rep.* **7**, 1111–1116 (2006).
46. Kim, S. J., Rahbar, R. & Hegde, R. S. Combinatorial control of prion protein biogenesis by the signal sequence and transmembrane domain. *J. Biol. Chem.* **276**, 26132–26140 (2001).
47. Hölscher, C., Bach, U. C. & Dobberstein, B. Prion protein contains a second endoplasmic reticulum targeting signal sequence located at its C terminus. *J. Biol. Chem.* **276**, 13388–13394 (2001).
48. Ott, C. M. & Lingappa, V. R. Signal sequences influence membrane integration of the prion protein. *Biochemistry* **43**, 11973–11982 (2004).
49. Hegde, R. *et al.* A transmembrane form of the prion protein in neurodegenerative diseases. *Science* **279**, 827–834 (1998).
50. Glatzel, M., Stoeck, K., Seeger, H., Luhrs, T. & Aguzzi, A. Human prion diseases: molecular and clinical aspects. *Arch. Neurol.* **62**, 545–552 (2005).
51. Gafvelin, G. & von Heijne, G. Topological 'frustration' in multi-spanning *E. coli* inner membrane proteins. *Cell* **77**, 401–412 (1994).
52. Ota, K., Sakaguchi, M., von Heijne, G., Hamasaki, N. & Mihara, K. Forced transmembrane orientation of hydrophilic polypeptide segments in multispanning membrane proteins. *Mol. Cell* **2**, 495–503 (1998).
53. Higy, M., Junne, T. & Spiess, M. Topogenesis of membrane proteins at the endoplasmic reticulum. *Biochemistry* **43**, 12716–12722 (2004).
54. Denzer, A. J., Nabholz, C. E. & Spiess, M. Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO J.* **14**, 6311–6317 (1995).
55. Goder, V., Bieri, C. & Spiess, M. Glycosylation can influence topogenesis of membrane proteins and reveals dynamic reorientation of nascent polypeptides within the translocon. *J. Cell Biol.* **147**, 257–266 (1999).
56. Nishiyama, K., Suzuki, T. & Tokuda, H. Inversion of the membrane topology of SecG coupled with SecA-sependent preprotein translocation. *Cell* **85**, 71–81 (1996).
57. Nagamori, S., Nishiyama, K. & Tokuda, H. Membrane topology inversion of SecG detected by labeling with a membrane-impermeable sulfhydryl reagent that causes a close association of SecG with SecA. *J. Biochem. (Tokyo)* **132**, 629–634 (2002).
58. van der Sluis, E. O., van der Vries, E., Berrelkamp, G., Nouwen, N. & Driessen, A. J. Topologically fixed SecG is fully functional. *J. Bacteriol.* **188**, 1188–1190 (2006).
59. Lu, Y. *et al.* Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol. Biol. Cell* **11**, 2973–2985 (2000).
Shows that transmembrane helices can reorientate across the membrane during translocon-mediated membrane-protein insertion.
60. Lambert, C. & Prange, R. Dual topology of the hepatitis B virus large envelope protein: determinants influencing post-translational pre-S translocation. *J. Biol. Chem.* **276**, 22265–22272 (2001).
61. Lambert, C., Mann, S. & Prange, R. Assessment of determinants affecting the dual topology of hepadnaviral large envelope proteins. *J. Gen. Virol.* **85**, 1221–1225 (2004).
62. Lambert, C. & Prange, R. Chaperone action in the posttranslational topological reorientation of the hepatitis B virus large envelope protein: implications for translocational regulation. *Proc. Natl Acad. Sci. USA* **100**, 5199–5204 (2003).
63. Graschopf, A. & Bläsi, U. Molecular function of the dual-start motif in the λ S holin. *Mol. Microbiol.* **33**, 569–582 (1999).
64. Cocquerel, L. *et al.* Topological changes in the transmembrane domains of hepatitis C virus envelope glycoproteins. *EMBO J.* **21**, 2893–2902 (2002).
65. Wang, X. Y., Bogdanov, M. & Dowhan, W. Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. *EMBO J.* **21**, 5673–5681 (2002).
66. Bogdanov, M., Heacock, P. N. & Dowhan, W. A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J.* **21**, 2107–2116 (2002).
Shows that changes in lipid composition can reversibly affect membrane-protein topology.
67. Zhang, W., Campbell, H. A., King, S. C. & Dowhan, W. Phospholipids as determinants of membrane protein topology. Phosphatidylethanolamine is required for the proper topological organization of the γ-aminobutyric acid permease (GabP) of *Escherichia coli*. *J. Biol. Chem.* **280**, 26032–26038 (2005).
68. Zhang, W., Bogdanov, M., Pi, J., Pittard, A. J. & Dowhan, W. Reversible topological organization within a polytopic membrane protein is governed by a change in membrane phospholipid composition. *J. Biol. Chem.* **278**, 50128–50135 (2003).

69. Xie, J., Bogdanov, M., Heacock, P. & Dowhan, W. Phosphatidylethanolamine and monoglucosyldiacyl glycerol are interchangeable in supporting topogenesis and function of the polytopic membrane protein lactose permease. *J. Biol. Chem.* **281**, 19172–19178 (2006).
70. van Klompenburg, W., Nilsson, I. M., von Heijne, G. & de Kruijff, B. Anionic phospholipids are determinants of membrane protein topology. *EMBO J.* **16**, 4261–4266 (1997).
71. Andersson, H. & von Heijne, G. Membrane protein topology: Effects of $\Delta\mu H^+$ on the translocation of charged residues explain the 'positive inside' rule. *EMBO J.* **13**, 2267–2272 (1994).
72. Goder, V., Junne, T. & Spiess, M. Sec61p contributes to signal sequence orientation according to the positive-inside rule. *Mol. Biol. Cell* **15**, 1470–1478 (2004).
The first demonstration that specific residues in Sec61α can influence membrane-protein topology.
73. Shimizu, T., Mitsuke, H., Noto, K. & Arai, M. Internal gene duplication in the evolution of prokaryotic transmembrane proteins. *J. Mol. Biol.* **339**, 1–15 (2004).
The first systematic survey of membrane-protein-topology evolution.
74. Murata, K. *et al.* Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
75. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).
76. Liu, Y., Gerstein, M. & Engelman, D. M. Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism. *Proc. Natl Acad. Sci. USA* **101**, 3495–3497 (2004).
77. Bernsel, A. & von Heijne, G. Improved membrane protein topology prediction by domain assignments. *Protein Sci.* **14**, 1723–1728 (2005).
78. von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **341**, 456–458 (1989).
79. Nilsson, I. M. & von Heijne, G. Fine-tuning the topology of a polytopic membrane protein. Role of positively and negatively charged residues. *Cell* **62**, 1135–1141 (1990).
80. Butler, P. J., Ubarretxena-Belandia, I., Warne, T. & Tate, C. G. The *Escherichia coli* multidrug transporter EmrE is a dimer in the detergent-solubilised state. *J. Mol. Biol.* **340**, 797–808 (2004).
81. Nishino, K. & Yamaguchi, A. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *J. Bacteriol.* **183**, 5803–5812 (2001).
82. Chen, G. Q., Cui, C., Mayer, M. L. & Gouaux, E. Functional characterization of a potassium-selective prokaryotic glutamate receptor. *Nature* **402**, 817–821 (1999).
83. Grigorieff, N., Ceska, T. A., Downing, K. H., Baldwin, J. M. & Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **259**, 393–421 (1996).
84. Bowie, U. J., Flip-flopping membrane proteins. *Nature Struct. Mol. Biol.* **13**, 94–96 (2006).

Acknowledgements
Gunnar von Heijne is supported by grants from the Swedish Foundation for Strategic Research, the Marianne and Marcus Wallenberg Foundation, the Swedish Cancer Foundation, the Swedish Research Council and the European Commission (BioSapiens).

Competing interests statement
The author declares no competing financial interests.

DATABASES

The following terms in this article are linked online to:
Pfam: http://www.sanger.ac.uk/Software/Pfam
DUF606
Protein Data Bank: http://www.rcsb.org/pdb
1KPK | 1T5S | 2BRD | 2F2M
UniProtKB: http://ca.expasy.org/sprot
CrcB | EmrE | PotE | SugE | YidC | Ydgl

FURTHER INFORMATION

Gunnar von Heijne's homepage:
http://www.cbr.su.se/content/view/64/41
Center for Biomembrane Research: http://www.cbr.su.se
Membrane Protein Resources:
http://blanco.biomol.uci.edu/MemPro_resources.html
Visual Molecular Dynamics software:
http://www.ks.uiuc.edu/Research/vmd

SUPPLEMENTARY INFORMATION

See online article: S1 (movie) | S2 (movie) | S3 (movie) | S4 (movie)
Access to this links box is available online.