Solving the membrane protein folding problem

James U. Bowie[1]

One of the great challenges for molecular biologists is to learn how a protein sequence defines its three-dimensional structure. For many years, the problem was even more difficult for membrane proteins because so little was known about what they looked like. The situation has improved markedly in recent years, and we now know over 90 unique structures. Our enhanced view of the structure universe, combined with an increasingly quantitative understanding of fold determination, engenders optimism that a solution to the folding problem for membrane proteins can be achieved.

In Ilse Aichinger's famous short story *The Bound Man*, a man awakes from a coma to find himself bound by ropes[1]. He learns to move gracefully within these constraints and eventually becomes a circus performer. Similarly, membrane proteins must perform complex signalling and transport functions within the strict confines of a lipid bilayer. This requires elegant structural adaptations to their environment.

Our early views of membrane-protein structure were largely shaped by the pioneering work of Henderson and colleagues, who established a modest resolution view of bacteriorhodopsin in 1975 (ref. 2). Confirming the prescient prediction of Lenard and Singer in 1966 (ref. 3), the structure revealed a bundle of long helical rods traversing the membrane. This suggested that membrane proteins could be largely thought of as an assembly of transmembrane (TM) helices. This view was reinforced by the stretches of hydrophobic residues found in membrane-protein sequences that were long enough to form membrane-spanning helices[4]. But recent structures have challenged this simple concept of membrane-protein architecture. In this review, I will briefly illustrate some of the surprising twists and turns that polypeptide chains make in the bilayer and then discuss our current understanding how these structures are built. Finally, I will argue that with concerted effort, practical solutions to the membrane-protein-folding problem are possible.

Complex architectures

Protein structure within the bilayer can be divided into two general types: β-barrels and bundles of α-helices. Because the folding problem for β-barrels and helix bundles is very different, and because β-barrel proteins are much less common[5], I will focus on the helix bundle class. Figure 1 shows our progress with helix bundle protein-structure determination since the first high-resolution membrane protein structure was solved 20 years ago (reviewed in ref. 6). A decade later, the pace of structure determination quickened. We now know 52 unique helix bundle structures, and 38 of these were solved in the past 5 years. As a result, we have a much clearer view of the structural diversity exhibited by membrane proteins. Although long TM helices are still considered a fundamental building block, polypeptide chains can be organized into other complex patterns.

Perhaps the first dramatic departure from the standard view was the structure of the glycerol/water channel GlpF, determined in 2000 (ref. 7). This structure illustrates some of the complexities of membrane-protein architecture. Two views of the GlpF monomer (the protein is a tetramer) are shown in Fig. 2, and I will call them front and back. The back side presents a simple picture of TM helices packed together roughly parallel to the membrane normal. Few would have been surprised by this image, even 30 years ago. The front side presents a more complex view, however. Most notable is the pair of helices that penetrate half way into the bilayer, where their amino-termini meet. Such half TM helices are not very common, but are also not particularly unusual (about 1 in 20 of all TM helices)[8]. Also visible in the front view is a common feature of membrane proteins: a highly distorted TM helix. About 60% of TM helices contain significant bends or other distortions[9]. Finally, loop conformations between

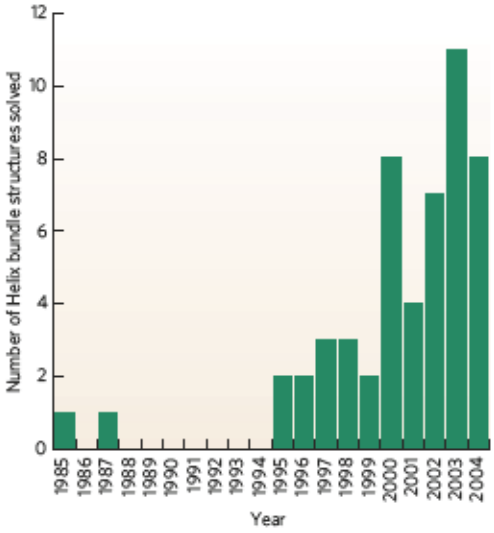

Figure 1 | Progress of helix bundle membrane protein structure determination. Only unique structures are included. The data were obtained from a website maintained by Stephen White (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html). For a similar plot including all membrane protein structures, see ref. 6.

[1]Department of Chemistry and Biochemistry, UCLA-DOE Center for Genomics and Proteomics, Molecular Biology Institute, Boyer Hall, UCLA, 611 Charles E. Young Drive E, Los Angeles, California 90095-1570 USA.

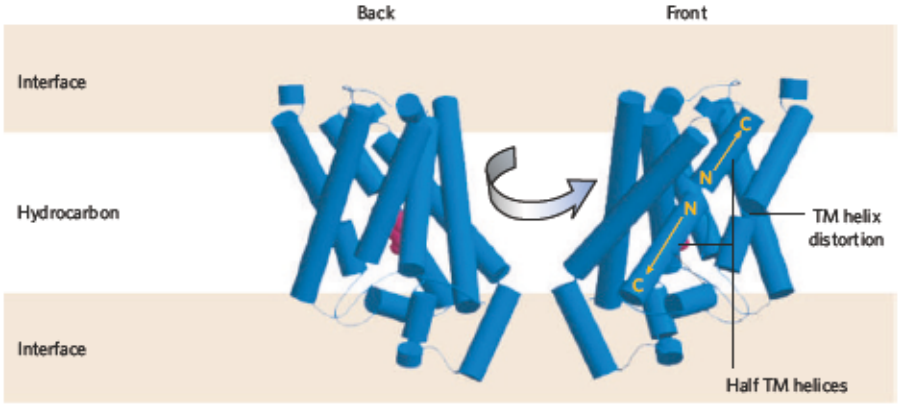

Figure 2 | The structure of the glycerol channel GlpF. Two views, labelled front and back, are shown[7]. They are rotated 180° with respect to each other along the membrane normal. The N and C termini of the half TM helices are labelled in the front view. A glycerol in the channel is shown in CPK mode. The structure was approximately positioned in the hydrocarbon core of the membrane by finding the most hydrophobic 30-Å slab of the protein perpendicular to the four-fold axis of the tetramer.

TM helices do not seem notably restricted by the membrane. Although packed TM helices remain by far the most common membrane-protein structural feature, protein chains are clearly not bound by such rigid limitations. Membrane proteins have developed tricks to get their chains to go where needed to satisfy functional imperatives. To solve the membrane-protein-folding problem we will need to understand the genesis of these structural idiosyncrasies.

Two fundamental stages in membrane-protein folding

As proposed originally by Popot and Engelman[10], it is convenient to break membrane-protein folding into two stages: insertion and folding. These stages are illustrated in Fig. 3.

In the first stage, the membrane protein is inserted across the bilayer. This stage can be both directed and catalysed by a translocon complex. In this manner, some of the membrane-inserted segments, and their topology, are established. The first phase marks membrane protein folding as fundamentally different from soluble protein folding because the insertion and topology decisions are made in consultation with the translocon complex[11,12]. By contrast, the fold of soluble proteins is normally defined entirely within the sequence itself[13]. To understand the first stage, we will need to learn how the sequence and the translocon communicate.

In the second stage, the tertiary and quaternary structures are built. This second stage involves assembly and reorientation of the TM segments established in the first phase, the additional insertion of re-entrant portions of the chain and subunit oligomerization. It may make sense to separate TM helix packing from other folding steps[14], but I lump these interrelated steps together in this review. A complete description of the second stage will require an understanding of amino-acid preferences for different portions of the bilayer, the energetics of interactions within the protein itself as well as interactions between the protein and the bilayer.

A thermodynamic view of insertion

Unlike soluble proteins, membrane proteins reside in a variable and anisotropic environment[1]. Figure 4 shows a snapshot from a molecular dynamics simulation of a fluid POPC (1-palmitoyl 2-oleoyl phosphatidylcholine) bilayer[16]. It highlights the conformational heterogeneity of the lipid conformations and variable molecular compositions in different regions[17]. Although somewhat arbitrary, the membrane is typically divided into two general regions. The hydrocarbon core in the centre is roughly 30 Å wide in a typical bilayer and is dominated by the aliphatic lipid chains. The interfacial region of the bilayer comprises lipid headgroups and considerable bound water. The interfacial region is quite large, roughly 15 Å across, and is therefore a major feature of a membrane protein's environment. The atom distributions shown in Fig. 4 illustrate how the properties of the bilayer change markedly along the direction of the bilayer normal. Starting from the centre, the environment changes from extremely apolar to highly polar/charged and finally to bulk water, all in a space of about 30 Å. Clearly, our descriptions of membrane-protein-folding energetics will need to vary as a function of bilayer depth.

A characteristic feature of helical membrane-protein sequences is their hydrophobic stretches, which are approximately 20 residues long[10]. A 20-residue segment is just long enough to span the hydrocarbon core of a typical bilayer in a helical conformation. Hydrophobicity is an important feature defining TM helices, which implies that thermodynamic partitioning between the water and the bilayer plays a significant role in membrane insertion and/or maintenance of the membrane protein in the bilayer. Figure 5 shows free energies for transferring each of the 20 amino acids from water into octanol, which serves as a model for the hydrocarbon core[15,18]. According to this scale, transferring the backbone (a glycine residue) is unfavourable by 1.25 kcal mol^{-1} per residue, arguing that side-chain hydrophobicity drives the equilibrium in favour of insertion. In other words, there must be a threshold of side-chain hydrophobicity for the segment to favour bilayer insertion[19]. For example, according to this scale, a 20-residue polyalanine sequence would not be sufficiently hydrophobic to partition into the hydrocarbon core ($\Delta G = 10$ kcal mol^{-1}), but replacing five

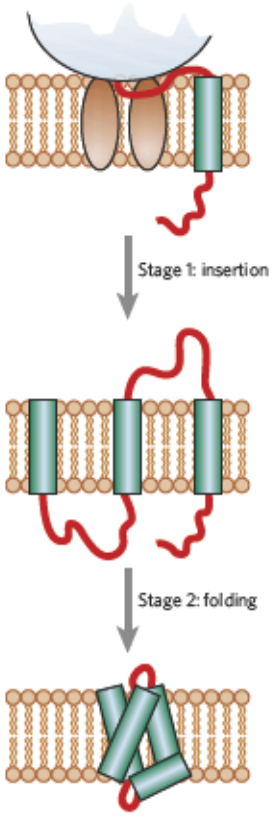

Figure 3 | Two stages of membrane protein folding[10]. See the text.

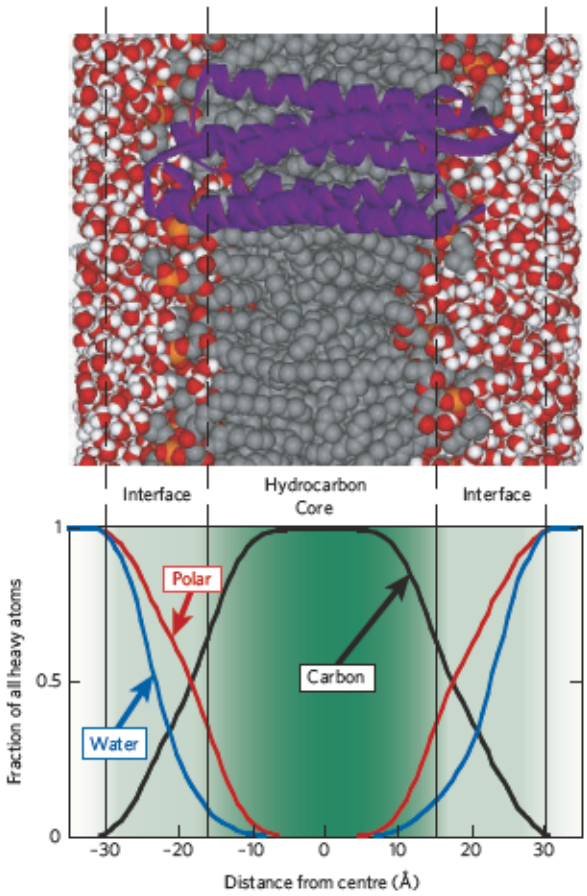

Figure 4 | A snapshot of a molecular dynamics simulation of a POPC bilayer[16]. The structure of bacteriorhodopsin is shown in front of the bilayer to give a sense of scale and was not included in the simulation. The graph at the bottom shows the fraction of carbon atoms (black line) the fraction of phosphorus and oxygen atoms generating polar moeties (red line) and the fraction of oxygen atoms from water (blue line). This figure was inspired by an earlier review by White and Wimley[80].

of the alanines with five leucines would be just favourable ($\Delta G = -1.25$ kcal mol^{-1}). Recent experiments on translocon-catalysed TM insertion indicate that this prediction is remarkably accurate[20]. Nevertheless, octanol is not a perfect model of the hydrocarbon core, which is not a homogeneous solvent, and transfer free energies vary as a function of bilayer depth[21]. As discussed below, new translocon-insertion experiments promise significant advancements over this basic thermodynamic model.

A biological view of insertion

Many membrane proteins are inserted as they emerge from the ribosome through a protein-conducting channel called the SecY complex in bacteria and the Sec61 complex in eukaryotes. The SecY/Sec61 channel must do a number of extraordinary things as a nascent polypeptide emerges from the ribosome. First, it needs to allow passage of the polypeptide chain through the membrane without leakage of ions or other small molecules. Second, it needs to decide whether to reverse the orientation of the emerging segment of the protein chain. Third, it needs to decide whether the emerging segment of the protein should be passed through to the aqueous compartment on the other side or slid into the membrane. Fourth, if it decides that the segment should go into the membrane, it needs to open the channel laterally and pass the segment into the bilayer, again without membrane leakage. A structure of the SecY/Sec61 translocon complex from *Methanococcus jannaschii* has now been determined[22], which markedly advances our ideas about how these complex functions can be accomplished.

The solved structure of the translocon complex, shown in Fig. 6, contains three polypeptide chains α, β and γ. The core of the complex is the large α subunit with ten TM helices surrounding the probable translocation pore. The structure is completed by the γ subunit, which closes off the α subunit on one end, and by the β subunit, which contributes an additional TM helix. The overall shape looking down from above or below the membrane resembles a clamshell that could possibly open on the side opposite the γ subunit (see Fig. 6). From the side, the channel is roughly hour-glass shaped and is accessible to water, but the aqueous channel is blocked by a short, helical plug from the α subunit. The structure solved is of the closed form of the channel without a translocating polypeptide. Nevertheless, the structure, coupled with considerable biochemical data, enabled Rapoport and colleagues[22,23] to suggest a plausible model for the general protein translocation process, which I will briefly summarize.

As the protein emerges from the ribosome, the plug slides out of the way, allowing the polypeptide to pass through the channel. The constriction in the channel is lined by hydrophobic residues (dark blue ring), which may provide a flexible self-sealing mechanism, maintaining a barrier between the two compartments (other mechanisms may operate as well). If the segment of the protein is destined for the external aqueous compartment, this *status quo* can be maintained. If the segment needs to be inserted into the membrane, however, something else needs to happen. The present model suggests that TM helices can be passed into the membrane by opening the channel to the side opposite the γ subunit (see Fig. 6).

The topology decision

How is the topology of the membrane-inserted segments established? Gunnar von Heijne and colleagues discovered that the cytoplasmic side of membrane proteins tends to be positively charged: the so-called positive-inside rule[24]. It is possible that part of this preference results from electrostatic potential differences between the two compartments and/or lipid composition differences between the two membrane leaflets. It is now clear, however, that the translocon itself plays an important role in the topology decision. Goder *et al.*[12] showed that by reversing the charge on positively and negatively charged residues in the *Saccharomyces cerevisiae* Sec61 α subunit, the topology of membrane proteins could be reversed. Thus, the same sequence can have a differ-

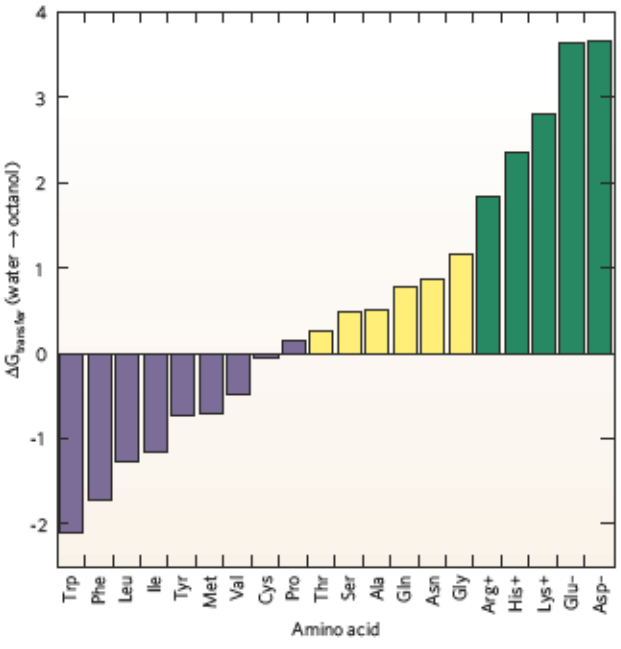

Figure 5 | Free energies for transfer of amino acids from water to octanol[15]. Charged residues are shown in green bars, polar residues in yellow bars and hydrophobic residues in purple bars.

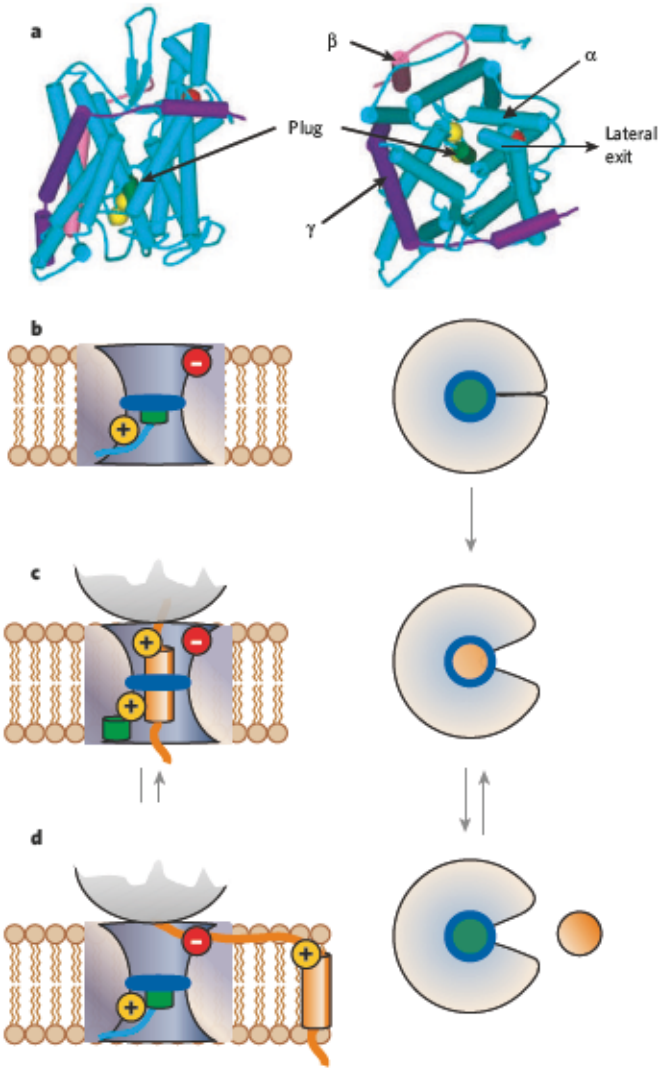

Figure 6 | The Sec61 translocon structure and mechanism model[23]. **a**, The α subunit is shown in light blue, the β subunit in purple and the γ subunit in pink. The plug helix that blocks the pore is shown in green. Key charged residues that help define the topology of translocating polypetides in *S. cerevisiae* are shown in yellow (positive charge) and red (negative charge). **b**, Schematic side and top views of the Sec61 translocon showing the locations of the charges, the plug helix in green and the hydrophobic collar shown by the dark blue ring. **c**, A schematic view of a the translocon with a nascent polypeptide (orange) emerging from the ribosome (grey). A positively charged residue helps define the topology of this segment as N-terminal first. **d**, A schematic view of the translocon opening the lateral gate so the helix can exit to the membrane if the partitioning is favourable.

ent topology depending on the sequence of the translocon. The positions of the key charged residues identified in yeast can be mapped onto the *M. jannashii* structure as shown in Fig. 6 (refs 11, 23). Because of these charged residues, as the emerging polypeptide chain approaches the region of the plug it would probably experience a strong positive electrostatic potential. If the emerging segment was positively charged, electrostatic repulsion could drive a reorientation of the polypeptide, pushing the N terminus away from the constriction and toward the negative charge at the top of the pore, thereby flipping the orientation of the chain. Other factors also contribute to the topology decision, such as synthesis rate and the length of the hydrophobic segment[12,23]. Moreover, the important charged residues identified in yeast are not conserved in all Sec61 translocons, suggesting either that altenative residues contribute to this mechanism or that other solutions to this problem operate in different organisms. This could be one of the reasons that it is so difficult to express membrane proteins in heterologous organisms.

The insertion decision

How does the translocon know which segments should be inserted into the membrane? Recent work by Hessa *et al.* has significantly improved our understanding of the insertion code[20]. They examined the probability that a 19-residue segment would be inserted into the endoplasmic reticulum membrane or passed through the translocon and into the aqueous lumen. As discussed above, physical chemistry suggests that a polyalanine sequence would not be sufficiently hydrophobic to insert into the membrane, but if five alanines were replaced with leucine, it would be just at the threshold for insertion. Remarkably, this is exactly what was observed. As the hydrophobicity of a polyalanine segment was increased by adding leucines, the insertion probability increased. Moreover, the insertion-probability increase could well be described by a Boltzmann distribution, as though the segments partition according to their transfer free energies. By substituting different amino acids into the test segment, it was possible to measure an apparent free energy of biological partition for each of the 20 amino acids. The biological scale correlates reasonably well with measured free energies for transferring amino-acid side chains from water to octanol (see above). Thus, the insertion probability behaves as though the translocon is measuring the free energy of partitioning between the aqueous and membrane phases.

How can the translocon detect the partitioning free energy? A plausible model is that the translocon frequently opens the lateral gate during protein translocation (see Fig. 6)[25]. When this happens, the nascent polypeptide segment in the translocon channel can sample both the aqueous and membrane phases. If the rate of synthesis is sufficiently slow relative to this sampling process, equilibrium can be established. If the equilibrium favours the bilayer, the segment becomes inserted. Otherwise it continues straight through the channel. In this manner, the translocon can act as a catalyst for the partitioning process. The structural details of this opening and how it can be accomplished without leakage is still something of a mystery.

The idea that the translocon catalyses partitioning between the aqueous and membrane phases creates a wonderful opportunity for refining our understanding of insertion thermodynamics beyond the model solvent studies discussed above. Because the membrane is not a homogeneous environment, insertion probability does not depend just on composition, but also on the location of individual residues within the TM segment. Insertion probabilities for polar residues are particularly dependent on their position in the TM segment[20]. These differences could be due to variations in the environment at different bilayer depths, as well as to compensating structural changes in both the bilayer and the protein. For example, long polar amino acids can arrange their side chains to move their polar atoms away from the hydrocarbon core in the direction of the increasing polarity gradient. This so-called 'snorkelling'[26] can make the placement of polar amino acids increasingly favourable as they approach the interfacial region. Translocon-insertion experiments suggest that it is about 2 kcal mol^{-1} more favourable to place an arginine residue near the edge of the hydrocarbon core than at the centre[27]. This factor is so significant that it is even possible to insert the S4 helix of a voltage-gated potassium channel, which contains four arginine residues, into a membrane[27]. Moving beyond simple consideration of segment hydrophobicity to define factors that contribute to insertion probability will greatly improve our ability to predict which segments of a protein become inserted. This is a key step on the way to solving the membrane-protein-folding problem.

Post-insertion folding

After (and probably during) insertion, the protein can begin a search for the stable, native conformation. We currently have little experimental knowledge of what the membrane-inserted unfolded state looks like or the pathway to the folded state in bilayers. Nevertheless, it is reasonable to surmise that a major component of the native-state conformational search will be packing of the stably inserted helices

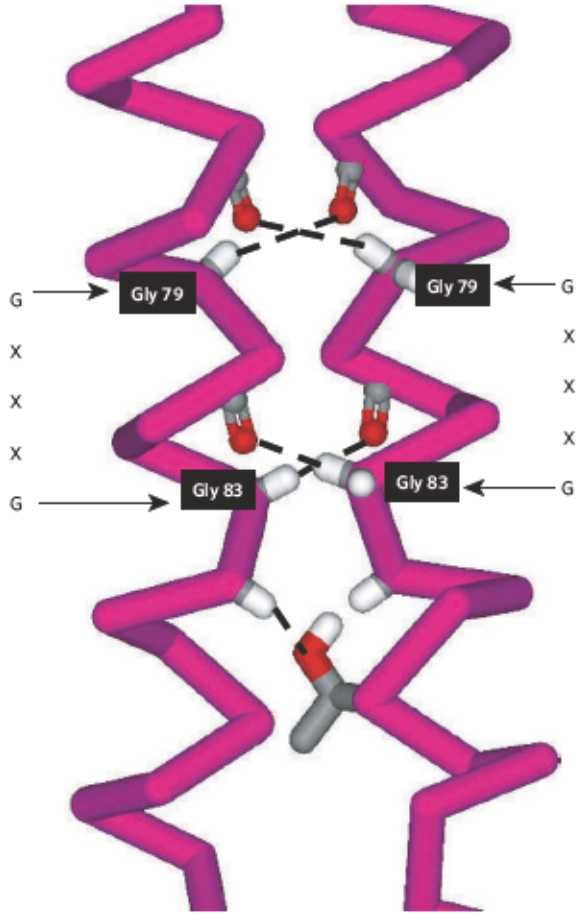

Figure 7 | The glycophorin A dimer[39]. The two key packing interface glycines found in the GXXXG sequence are highlighted. A network of Cα-H···O hydrogen bonds between backbone carbonyls and the Cα hydrogens of the glycine residues are shown, along with a Cα-H···O hydrogen bond made by a threonine side chain. In the atoms shown, carbon is grey, oxygen is red and hydrogen is white. They hydrogen bonds are indicated by the dashed lines.

from stage 1 (refs 10, 14). Considerably more needs to happen, however, to attain a complex structure such as GlpF (Fig. 2). Additional segments of the protein chain must be inserted into the membrane or at least rearranged within the membrane. We also cannot rule out the possibility of topology rearrangements during folding. Some membrane proteins can spontaneously insert into bilayers[28], and Dowhan and co-workers have found that the topology of improperly inserted membrane proteins can self-correct if the lipid composition is altered after incorrect insertion, indicating that hydrophilic loop regions can be dragged through the bilayer[29].

Like the soluble-protein-folding problem, folding of membrane proteins probably proceeds down a funnel-shaped energy landscape to an energy minimum[30]. Consistent with a folding funnel view is the observation of multiple pathways in the folding of bacteriorhodopsin[31]. A major difference between soluble protein folding and membrane protein folding, however, is that the starting point is much more constrained, because much of the secondary structure and topology will be set by the insertion process. Thus, the unfolded protein is much farther down the folding funnel and closer to the folded state compared with soluble proteins. It seems clear that the folding-energy landscape is defined by a complex interplay between various forces, including polypeptide partitioning in the bilayer (discussed above), interactions between lipid and protein and interactions within the protein itself. I will briefly discuss these factors in the following sections. I will focus on forces acting within the bilayer because the extensive work on water-soluble protein folding will largely be applicable to chain folding in the water-exposed portions.

Basic features of transmembrane helix packing

As a consequence of environmental constraints, TM helices exhibit a narrower distribution of packing angles than found in soluble proteins, with a strong preference at around 20° (ref. 32). In fact, more than 60% of all TM-helix-packing angles fall in the range of 0° to 40°. Part of this preference is due to the favourable orientation of TM helices along the membrane normal, which tends to favour small packing angles[33]. The 20° packing angle is particularly favourable because it facilitates interhelix side-chain interdigitation[34]. TM helices can also exhibit superhelical wrapping[35], which limits the helix divergence that would occur if they remained perfectly straight.

TM-helix-packing interfaces show a preference for small residues[36]. A well-known example is the strong TM helix dimer from glycophorin A shown in Fig. 7. The core of this interface, a GXXXG sequence motif, is found in many helix oligomers[37]. The glycine residues in this motif face each other in the dimer, allowing close packing of the helices. There are several possible explanations for the small-residue preference. First, small residues allow close approach of the helices, leading to improved packing[38]. Second, there is a lower entropy cost for fixing the conformation of small side chains in the folded protein compared with larger residues[39]. Finally, small side chains expose polar backbone atoms that could lead to favourable interactions[40,41] (see below). As a measure of the significance of this preference, a simple scoring function that considers only residue size is surprisingly effective for packing TM helices[42].

Deviant helices

TM helices contain many local distortions including kinks and short stretches of π or 3_{10} helices, leading to a deviation of the helix axis and/or a shift in the regularity of side-chain positions[43,44]. Given their prevalence, predicting the location of helix distortions will be an important piece of the protein-folding puzzle. There are now encouraging signs. For example, Yohannan *et al.* found that kinks could be identified with more than 90% reliability simply by looking at proline abundance in a multiple sequence alignment[9]. Rigoutsos *et al.* report the identification of sequence patterns predictive of helix distortions[45], and local sequence features have also been suggested to alter kink magnitude[46].

Why are TM helices so frequently kinked? One probable reason is that kinks enable the small structural adjustments needed to position functional groups precisely, which could facilitate functional diversification of a common architecture[9]. For example, G-protein-coupled receptors all have seven TM helices but respond to a remarkable range of signals. This would probably be impossible if their construction were limited to seven rigid rods. Another possibility is that the helical distortions provide weak points in the helical rods that facilitate movements needed for protein function[44]. Finally, proline residues were found to block off-pathway β-sheet formation during CFTR (cystic fibrosis transmembrane conductance regulator) folding[47]. Thus, some proline residues can be important for directing folding toward the native state structure.

Internal forces stabilizing membrane proteins

TM regions are dominated by apolar side chains, and van der Waals interactions play an important role in stabilizing interactions in the hydrocarbon core region of the bilayer. For example, the interface of the synaptobrevin helix dimer seems to be constructed entirely from apolar side chains[48]. On the basis of the stability effects of side-chain deletions in bacteriorhodopsin and glycophorin A, side-chain burial contributes approximately 27 cal mol^{-1} $Å^{-2}$ (in detergent), which is similar to the contribution observed in soluble proteins[49]. In soluble proteins, a significant fraction of the energetic contribution of residue burial arises from the hydrophobic effect, which is not a dominant factor in the hydrocarbon core region of the bilayer. Somehow membrane proteins seem to make up for the loss of the hydrophobic effect. One possible factor is that packing is better in membrane proteins than soluble proteins, but there is no consensus on this point. Although by some measures membrane proteins are better packed than in soluble

proteins[50,51], others argue that membrane proteins contain more packing defects[52,53]. It is possible that locally well-packed regions of structure are particularly important and more than compensate for the packing defects observed elsewhere. Alternatively, the unfolded state of membrane proteins may pack less well with the bilayer than soluble proteins pack with water. Although the packing density of water is lower (0.36; ref. 54) than lipids in bilayers (about 0.42 in the centre and about 0.65 near headgroups)[55], water is freer to pack around the protein side chains than are lipids.

Although apolar residues dominate in bilayer-embedded regions of membrane proteins, the interiors of membrane proteins are more polar than their exteriors[56] and there is ample potential for hydrogen-bonding interactions[57]. In aqueous solution, the strength of hydrogen bonds is diminished by the high dielectric effect in water and competition from water for hydrogen bonds. In the apolar hydrocarbon core, however, hydrogen bonds have the potential to be very strong[15,58], worth possibly more than 5 kcal mol^{-1} (ref. 15). Experimental measurements of hydrogen-bond contributions between TM helices are significantly lower than that, however, and quite variable with an upper limit in the range of 2 kcal mol^{-1} (refs 49, 59).

Several factors can reduce hydrogen-bond contributions. First, the position in the bilayer seems to be very important[59]. If placed near the edge of the hydrocarbon core, polar moieties from the interfacial region can make competing hydrogen bonds in the unfolded state, reducing the net contribution to stability[59]. Second, it may be difficult to arrange good hydrogen-bond geometries given the constraints of the polypeptide chain in a folded structure[49,60]. Nevertheless, appropriately placed hydrogen bonds with good geometry can be quite significant. For example, the introduction of a single polar residue into an apolar TM helix can strongly drive oligomerization[59,60–64]. Networks of polar interactions have been described that probably play important roles in defining TM helix structure[57].

Because of the potential for forming strong hydrogen-bonding interactions, polar residues must be used judiciously in membrane proteins or they could engage in inappropriate interactions, particularly in the crowded membrane environment[63]. Indeed, polar substitutions are the most common disease-causing mutations in membrane proteins[65]. For example, one of the earliest identified mutations in an oncogene was the valine to glutamic acid mutation in the TM domain of the Her2/Neu receptor. The glutamic acid forms a hydrogen bond in a Neu TM domain peptide[66] and the mutation seems to induce receptor aggregation and consequent activation[67]. In the CFTR chloride channel, a valine to asparagine mutation also seems to alter TM helix interactions through inappropriate hydrogen bonding[68].

In addition to strongly polar interactions, weakly polar interactions probably play a significant role in stabilizing membrane-protein structure[58]. Of particular note are Cα–H···O hydrogen bonds[40]. Interest in these carbon–hydrogen bonds was piqued by *ab initio* quantum mechanics calculations suggesting that Cα–H···O hydrogen bonds could be approximately half the strength of a traditional hydrogen bond[69]. Senes *et al.* demonstrated that they are also prevalent in membrane-protein structures[40]. An example of a striking network of potential Cα–H···O hydrogen bonds is found in the strong TM helix dimer of glycophorin A, shown in Fig. 7, which is made possible by the small glycine residues in the interface noted above.

So how strong are these carbon–hydrogen bonds? Like traditional hydrogen bonds, it seems that the answer depends on the context. In an elegant test, Arbely and Arkin found that hydrogen bonding altered the Cα–H bond stretching frequency and estimated a contribution of 0.9 kcal mol^{-1} (ref. 41). This may be an overestimate as it is based on *in vacuo* model data. Nevertheless, these results indicate that the interaction is favourable. By contrast, Yohannan *et al.* found that a Cα–H···O hydrogen bond in bacteriorhodopsin made little contribution to stability[70]. Mottamal and Lazaridis suggest that the difference in these results stems from the different hydrogen-bond geometries[71]. Thus, like normal hydrogen bonds, the structural details will probably define Cα–H···O hydrogen-bond strength. Because of the importance of hydrogen bonding in the membrane interior, a detailed understanding of factors that alter the strengths of hydrogen bonds will be an important step toward a solution to the membrane-protein-folding problem.

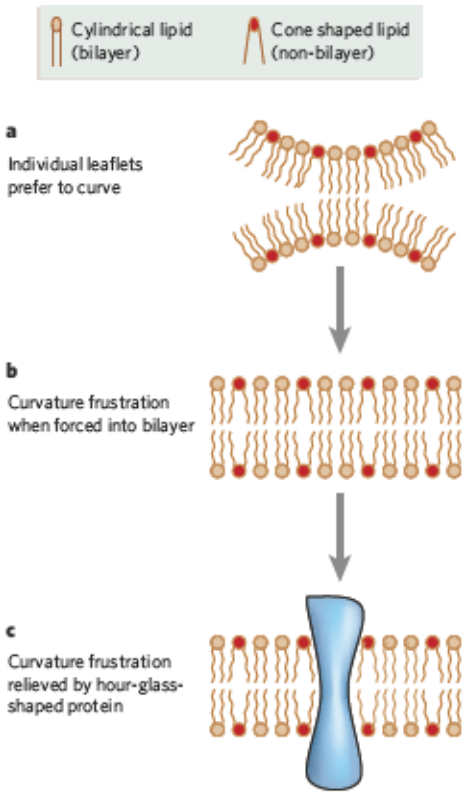

Figure 8 | Curvature elastic energy. Two types of lipid are depicted as either cylindrical or cone-shaped. **a**, The presence of a cone-shaped lipid in a bilayer will cause the two leaflets to want to curve away from one another. **b**, Forcing them into a bilayer causes overpacking in the hydrocarbon tails. **c**, An hour-glass-shaped protein can release some of this stored curvature elastic energy.

Non-specific driving forces in the bilayer

In addition to polarity gradients affecting side-chain partitioning discussed earlier, other bilayer properties can play a major role in membrane-protein folding, stability and structure[72]. Properties of specific lipids are treated in detail in an accompanying review. Here we will simply consider general effects. Although it is more correct to describe lipid packing in terms of intermolecular forces[72], in qualitative terms it is useful to think of different lipids as having distinct headgroup and hydrocarbon tail sizes (see Fig. 8). To form a bilayer, the lipid should have a cylindrical shape with the radius of the headgroup approximately matching the radius of the hydrocarbon tails. In some lipids, the hydrocarbon tails splay out so they cannot form a bilayer by themselves. Such 'non-bilayer' lipids can be incorporated into a bilayer if they are mixed with bilayer-forming lipids. As shown in Fig. 8, the shape mismatch creates a driving force for the bilayer leaflets to curve away from one another. Forcing them into a bilayer creates a curvature frustration[73], which can be simplistically described as an overpacking in the hydrocarbon region and an underpacking in the headgroup region.

The drive to relieve curvature elastic energy can shape membrane protein structure. For example, as shown in Fig. 8, an hour-glass shape should relieve the curvature strain. Indeed, Tamm and co-workers showed that curvature elastic energy has a significant effect on the stability of OmpA, a hour-glass-shaped β-barrel protein[74]. Also, Booth and co-workers have demonstrated that changes in lipid composition

predicted to change lateral packing pressure alter the folding pathway of bacteriorhodopsin[75]. The lack of curvature elastic energy in detergent micelles could be a factor in destabilizing or altering the structure of membrane proteins in detergent solution.

Different lipid compositions can alter the natural bilayer thickness. If the hydrophobic surface of the membrane protein is thicker or thinner than the hydrocarbon core of the bilayer, a tension is created because of the drive to shield hydrophobic surfaces from water. As shown in Fig. 9, this hydrophobic mismatch tension can be resolved in several ways. Either the protein can adjust to match the bilayer or the bilayer can deform to match the protein. *In vitro* experiments in purified systems demonstrate that in β-barrel proteins, the lipid deforms to match the protein[74,76]. By contrast, helical membrane proteins, which may be less rigid than β-barrel proteins, seem able to adjust to the hydrophobic thickness of the bilayer[77,78]. However, recent experiments with membranes isolated from cells that are filled with helical proteins suggest that membrane thickness is set by the proteins, not the lipids[79]. One possible explanation for this difference is the complex lipid composition in natural membranes that could facilitate the bilayer deformations necessary for hydrophobic matching (see Fig. 9). Indeed, curvature elastic strain does seem to facilitate hydrophobic matching of the β-barrel protein OmpA[74].

Specific lipid interactions

In addition to general bilayer properties, bound lipids have been seen in many crystal structures[72] and are often crucial for protein function. For example, cytochrome C oxidase is inactivated by the removal of cardiolipin[80] and KcsA requires anionic phospholipids[81]. Thus, lipids can act as cofactors for some membrane proteins and stabilize their structures.

Toward a solution to the membrane-protein-folding problem

A solution to the membrane protein-folding problem will probably follow the two stages outlined at the beginning of this review. We must first be able to accurately predict the location and topology of the initially inserted segments and then learn to fold the chain from this starting point. It seems likely that we will first have a practical solution for stage 1. Although much work remains, TM-helix prediction methods are already quite accurate at finding TM helices, if not their endpoints[8]. New experiments allowing us to effectively query the insertion code[20] along with the increasing structure database will improve these methods further. Moreover, perfection in first-stage prediction may not be necessary because it should be possible to start second-stage folding with a collection of topology models.

Can we solve the second-stage folding problem? In considering this question, it is useful to make a comparison with soluble protein folding. Although there has been limited practical success in *de novo* soluble protein folding, the best results are obtained with low-contact-order (many interactions close in sequence) helical proteins, that is, those that are similar to helix bundle membrane proteins[82]. With membrane proteins, we can add significant additional constraints to folding algorithms, however. In particular, we can start from a first-stage model that is closer to the final fold than the starting point for soluble protein folding. Moreover, the bilayer can be used to great advantage for structure-prediction efforts because it provides spatial information. In membrane proteins, second-stage folding can be directed not only by intra-protein interactions but also by bilayer position. For example, a trial conformation with a highly charged segment in the middle of the bilayer can be quickly rejected in the absence of mitigating factors. If soluble-protein folders could add this type of spatial information, it seems likely that robust prediction methods would be at hand. Thus, it is reasonable to think that reliable prediction methods for membrane proteins could be developed if the same level of effort were directed toward this goal as has been applied to soluble proteins. Indeed, practical structure prediction of TM-helix packing has already been possible in simple systems with the addition of limited experimental constraints[48,83–87]. To fold proteins with complexities such as half TM helices we will need to move beyond TM-helix packing algorithms and into global folding algorithms that can capture all the forces that drive such structural oddities. Ultimately, this will require a sophisticated description of the complex balance of forces that operate at different bilayer depths.

Our understanding of membrane-protein-folding determinants is still rudimentary and our database of structures is tiny in comparison with that of soluble proteins. But like the audience in Ilse Aichinger's imagined *Bound Man* witnessing the amazing physical adaptations of the man bound by ropes, we are beginning to comprehend how proteins adjust to limitations of the membrane. As indicated in this review, we are developing an increasingly quantitative understanding of the energetic forces that operate in a bilayer, and the pace of structure determination is accelerating. With the conformational restrictions imposed on membrane proteins, we are already proceeding more rapidly to effective structure prediction than has been possible with soluble proteins. Given the technical difficulties in obtaining membrane-protein structures, this is an area where prediction methods would be particularly welcome. ■

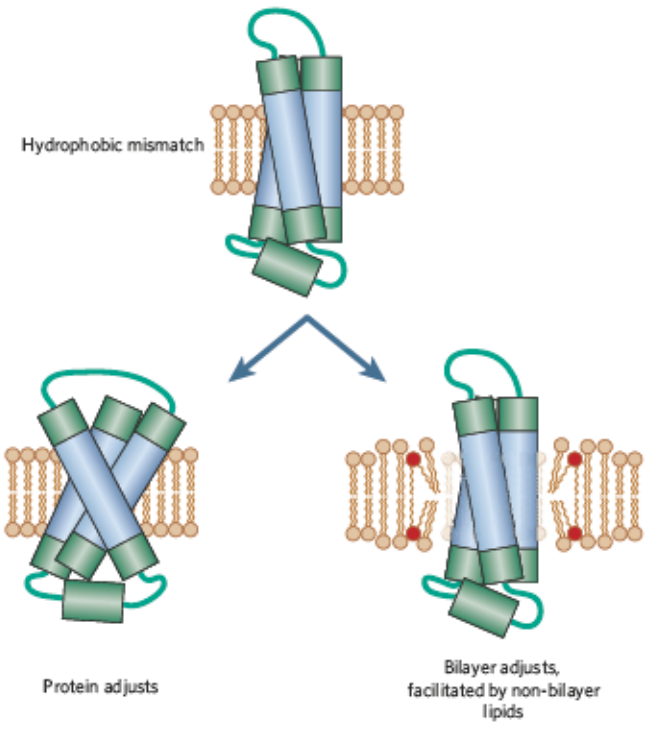

Figure 9 | Hydrophobic mismatch. If the hydrophobic region (blue) of a protein is thicker than the bilayer hydrocarbon core, either the protein can thin or the bilayer can thicken. Bilayer adjustment can be facilitated by non-bilayer lipids.

1. Aichinger, I. in *The Art of the Tale* (ed. Halpern, D.) 9–17 (Penguin Books, New York, NY, 1956).
2. Henderson, R. & Unwin, P. N. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28–32 (1975).
3. Lenard, J. & Singer, S. Protein conformation in cell membrane preparations as studied by optical rotatory dispersion and circular dichroism. *Proc. Natl Acad. Sci. USA* **56**, 1828–1835 (1966).
4. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
5. Bigelow, H. R., Petrey, D. S., Liu, J., Przybylski, D. & Rost, B. Predicting transmembrane beta-barrels in proteomes. *Nucleic Acids Res.* **32**, 2566–2577 (2004).
6. White, S. H. The progress of membrane protein structure determination. *Protein Sci.* **13**, 1948–1949 (2004).
7. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).

8. Cuthbertson, J. M., Doyle, D. A. & Sansom, M. S. Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng. Des. Sel.* **18,** 295–308 (2005).
9. 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).
10. Popot, J. & Engelman, D. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29,** 4031–4037 (1990).
11. White, S. H. & von Heijne, G. The machinery of membrane protein assembly. *Curr. Opin. Struct. Biol.* **14,** 397–404 (2004).
12. 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).
13. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181,** 223–230 (1973).
14. Engelman, D. M. *et al.* Membrane protein folding: beyond the two stage model. *FEBS Lett.* **555,** 122–125 (2003).
15. White, S. H. & Wimley, W. C. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 319–365 (1999).
16. Tieleman, D. P., Sansom, M. S. & Berendsen, H. J. Alamethicin helices in a bilayer and in solution: molecular dynamics simulations. *Biophys. J.* **76,** 40–49 (1999).
17. Wiener, M. C. & White, S. H. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of X-ray and neutron diffraction data. III. Complete structure. *Biophys J.* **61,** 437–447 (1992).
18. Wimley, W. C., Creamer, T. P. & White, S. H. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35,** 5109–5124 (1996).
19. Liu, L. P., Li, S. C., Goto, N. K. & Deber, C. M. Threshold hydrophobicity dictates helical conformations of peptides in membrane environments. *Biopolymers* **39,** 465–470 (1996).
20. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).
21. Wimley, W. C. & White, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Struct. Biol.* **3,** 842–848 (1996).
22. Van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427,** 36–44 (2004).
23. 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).
24. von Heijne, G. The distribution of positively charged reisues in bacterial inner membrane protiens correlates with the transmembrane topology. *EMBO J.* **5,** 3021–3027 (1986).
25. Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102,** 233–244 (2000).
26. Chamberlain, A. K., Lee, Y., Kim, S. & Bowie, J. U. Snorkeling preferences foster an amino acid composition bias in transmembrane helices. *J. Mol. Biol.* **339,** 471–479 (2004).
27. Hessa, T., White, S. H. & von Heijne, G. Membrane insertion of a potassium-channel voltage sensor. *Science* **307,** 1427 (2005).
28. Nagy, J. K., Lonzer, W. L. & Sanders, C. R. Kinetic study of folding and misfolding of diacylglycerol kinase in model membranes. *Biochemistry* **40,** 8971–8980 (2001).
29. 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 gamma-aminobutyric acid permease (GabP) of *Escherichia coli*. *J. Biol. Chem.* **280,** 26032–26038 (2005).
30. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4,** 10–19 (1997).
31. Lu, H. & Booth, P. J. The final stages of folding of the membrane protein bacteriorhodopsin occur by kinetically indistinguishable parallel folding paths that are mediated by pH. *J. Mol. Biol.* **299,** 233–243 (2000).
32. Bowie, J. Helix packing in membrane proteins. *J. Mol. Biol.* **272,** 780–789 (1997).
33. Huschilt, J., Millman, B. & Davis, J. Orientation of alpha-helical peptides in a lipid bilayer. *Biochim. Biophys. Acta* **979,** 139–141 (1989).
34. Chothia, C., Levitt, M. & Richardson, D. Helix to helix packing in proteins. *J. Mol. Biol.* **145,** 215–250 (1981).
35. Langosch, D. & Heringa, J. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins* **31,** 150–159 (1998).
36. Jiang, S. & Vakser, I. A. Side chains in transmembrane helices are shorter at helix-helix interfaces. *Proteins* **40,** 429–435 (2000).
37. Curran, A. R. & Engelman, D. M. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* **13,** 412–417 (2003).
38. Jiang, S. & Vakser, I. A. Shorter side chains optimize helix-helix packing. *Protein Sci.* **13,** 1426–1429 (2004).
39. MacKenzie, K. & Engelman, D. Structure-based prediction of the stability of transmembrane helix–helix interactions: the sequence dependence of glycophorin A dimerization. *Proc. Natl Acad. Sci. USA* **95,** 3583–3590 (1998).
40. Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. The Cα—H.O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl Acad. Sci. USA* **98,** 9056–9061 (2001).
41. Arbely, E. & Arkin, I. T. Experimental measurement of the strength of a C_{alpha}-H.O bond in a lipid bilayer. *J. Am. Chem. Soc* **126,** 5362–5363 (2004).
42. Fleishman, S. J. & Ben-Tal, N. A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane alpha-helices. *J. Mol. Biol.* **321,** 363–378 (2002).
43. Riek, R. P., Rigoutsos, I., Novotny, J. & Graham, R. M. Non-alpha-helical elements modulate polytopic membrane protein architecture. *J. Mol. Biol.* **306,** 349–362 (2001).
44. Cordes, F. S., Bright, J. N. & Sansom, M. S. Proline-induced distortions of transmembrane helices. *J. Mol. Biol.* **323,** 951–960 (2002).
45. Rigoutsos, I., Riek, P., Graham, R. M. & Novotny, J. Structural details (kinks and non-alpha conformations) in transmembrane helices are intrahelically determined and can be predicted by sequence pattern descriptors. *Nucleic Acids Res.* **31,** 4625–4631 (2003).
46. Deupi, X. *et al.* Ser and Thr residues modulate the conformation of pro-kinked transmembrane alpha-helices. *Biophys. J.* **86,** 105–115 (2004).
47. Wigley, W. C. *et al.* A protein sequence that can encode native structure by disfavoring alternate conformations. *Nature Struct. Biol.* **9,** 381–388 (2002).
48. Fleming, K. G. & Engelman, D. M. Computation and mutagenesis suggest a right-handed structure for the synaptobrevin transmembrane dimer. *Proteins* **45,** 313–317 (2001).
49. Faham, S. *et al.* Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol.* **335,** 297–305 (2004).
50. Eilers, M., Patel, A. B., Liu, W. & Smith, S. O. Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys. J.* **82,** 2720–2736 (2002).
51. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O. & Fleming, P. J. Internal packing of helical membrane proteins. *Proc. Natl Acad. Sci. USA* **97,** 5796–5801 (2000).
52. Hildebrand, P. W., Rother, K., Goede, A., Preissner, R. & Frommel, C. Molecular packing and packing defects in helical membrane proteins. *Biophys. J.* **88,** 1970–1977 (2005).
53. Adamian, L. & Liang, J. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* **311,** 891–907 (2001).
54. Richards, F. M. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82,** 1–14 (1974).
55. Marrink, S., Sok, R. & Berendsen, H. J. Free volume properties of a simulated lipid membrane. *J. Chem. Phys.* **104,** 9090–9099 (1996).
56. Rees, D., DeAntonio, L. & Eisenberg, D. Hydrophobic organization of membrane proteins. *Science* **245,** 510–513 (1989).
57. Adamian, L. & Liang, J. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* **47,** 209–218 (2002).
58. Chamberlain, A. K., Faham, S., Yohannan, S. & Bowie, J. U. Construction of helix-bundle membrane proteins. *Adv. Protein Chem.* **63,** 19–46 (2003).
59. Lear, J. D., Gratkowski, H., Adamian, L., Liang, J. & DeGrado, W. F. Position-dependence of stabilizing polar interactions of asparagine in transmembrane helical bundles. *Biochemistry* **42,** 6400–6407 (2003).
60. Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl Acad. Sci. USA* **98,** 2250–2255 (2001).
61. Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. Asparagine-mediated self-association of a model transmembrane helix. *Nature Struct. Biol.* **7,** 161–166 (2000).
62. Li, R. *et al.* Activation of integrin αIIbβ3 by modulation of transmembrane helix associations. *Science* **300,** 795–798 (2003).
63. Gratkowski, H., Lear, J. D. & DeGrado, W. F. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl Acad. Sci. USA* **98,** 880–885 (2001).
64. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nature Struct. Biol.* **7,** 154–160 (2000).
65. Partridge, A. W., Therien, A. G. & Deber, C. M. Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. *Proteins* **54,** 648–656 (2004).
66. Smith, S., Smith, C. & Bormann, B. Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbB-2 receptor. *Nature Struct. Biol.* **3,** 252–258 (1996).
67. Weiner, D., Liu, J., Cohen, J., Williams, W. V. & Green, M. A point mutation in the neu oncogene mimics ligand induction of receptor aggregation. *Nature* **339,** 230–231 (1989).
68. Therien, A. G., Grant, F. E. & Deber, C. M. Interhelical hydrogen bonds in the CFTR membrane domain. *Nature Struct. Biol.* **8,** 597–601 (2001).
69. Vargas, R., Garza, J., Dixon, D. A. & Hay, B. P. How strong is the Cα-H—O=C hydrogen bond? *J. Am. Chem. Soc.* **122,** 4750–4755 (2000).
70. Yohannan, S. *et al.* A Cα-H.O hydrogen bond in a membrane protein is not stabilizing. *J. Am. Chem. Soc.* **126,** 2284–2285 (2004).
71. Mottamal, M. & Lazaridis, T. The contribution of Cα-H.O hydrogen bonds to membrane protein stability depends on the position of the amide. *Biochemistry* **44,** 1607–1613 (2005).
72. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta* **1666,** 62–87 (2004).
73. Gruner, S. M. Intrinsic curvature hypothesis for biomembrane lipid composition: a role for nonbilayer lipids. *Proc. Natl Acad. Sci. USA* **82,** 3665–3669 (1985).
74. Hong, H. & Tamm, L. K. Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc. Natl Acad. Sci. USA* **101,** 4065–4070 (2004).
75. Allen, S. J., Curran, A. R., Templer, R. H., Meijberg, W. & Booth, P. J. Controlling the folding efficiency of an integral membrane protein. *J. Mol. Biol.* **342,** 1293–1304 (2004).
76. O'Keeffe, A. H., East, J. M. & Lee, A. G. Selectivity in lipid binding to the bacterial outer membrane protein OmpF. *Biophys. J.* **79,** 2066–2074 (2000).
77. Williamson, I. M., Alvis, S. J., East, J. M. & Lee, A. G. Interactions of phospholipids with the potassium channel KcsA. *Biophys. J.* **83,** 2026–2038 (2002).
78. Powl, A. M., East, J. M. & Lee, A. G. Lipid-protein interactions studied by introduction of a tryptophan residue: the mechanosensitive channel MscL. *Biochemistry* **42,** 14306–14317 (2003).
79. Mitra, K., Ubarretxena-Belandia, I., Taguchi, T., Warren, G. & Engelman, D. M. Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol. *Proc. Natl Acad. Sci. USA* **101,** 4083–4088 (2004).
80. Abramovitch, D. A., Marsh, D. & Powell, G. L. Activation of beef-heart cytochrome c oxidase by cardiolipin and analogues of cardiolipin. *Biochim. Biophys. Acta* **1020,** 34–42 (1990).
81. Valiyaveetil, F. I., Zhou, Y. & MacKinnon, R. Lipids in the structure, folding, and function of

the KcsA K^+ channel. *Biochemistry* **41,** 10771-10777 (2002).
82. Schonbrun, J., Wedemeyer, W. J. & Baker, D. Protein structure prediction in 2002. *Curr. Opin. Struct. Biol.* **12,** 348–354 (2002).
83. Briggs, J. A., Torres, J. & Arkin, I. T. A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins* **44,** 370–375 (2001).
84. Kim, S., Chamberlain, A. K. & Bowie, J. U. Membrane channel structure of Helicobacter pylori vacuolating toxin: role of multiple GXXXG motifs in cylindrical channels. *Proc. Natl Acad. Sci. USA* **101,** 5988–5991 (2004).
85. Sale, K., Faulon, J. L., Gray, G. A., Schoeniger, J. S. & Young, M. M. Optimal bundling of transmembrane helices using sparse distance constraints. *Protein Sci.* **13,** 2613–2627 (2004).
86. Adams, P., Engelman, D. & Brünger, A. Improved prediction fo the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins Struct. Func. Genet.* **26,** 257–261 (1996).
87. Fleishman, S. J., Unger, V. M., Yeager, M. & Ben-Tal, N. A Cα model for the transmembrane alpha helices of gap junction intercellular channels. *Mol. Cell* **15,** 879–888 (2004).
88. White, S. H. & Wimley, W. C. Hydrophobic interactions of peptides with membrane interfaces. *Biochim. Biophys. Acta* **1376,** 339–352. (1998).
89. MacKenzie, K., Prestegard, J. & Engelman, D. A transmembrane helix dimer: structure and implications. *Science* **276,** 131–133 (1997).

Acknowledgements I would like to thank members of my lab for helpful comments and NIH for support. J.U.B. is a Leukemia and Lymphoma Society Scholar.

Author Information Reprints and permissions is available at npg.nature.com/reprintsand permissions. The authors declare no competing financial interests. Correspondence should be addressed to J.U.B. (bowie@mbi.ucla.edu).