# Experimental and computational studies of determinants of membrane-protein folding

Jie Liang

Recent experiments and analysis have demonstrated the important roles of hydrogen bonding and polar–polar interactions in driving transmembrane helix–helix association. Further details of the energetics of helix–helix association and interhelical packing geometry are being mapped out. Many sequence motifs and a few spatial motifs promoting helical associations have also been identified.

**Addresses**
Department of Bioengineering, University of Illinois at Chicago, SEO 218, MC-063 851, South Morgan Street, Chicago, IL 60607, USA; e-mail: jliang@uic.edu

**Current Opinion in Chemical Biology** 2002, **6**:878–884

1367-5931/02/$ – see front matter
© 2002 Elsevier Science Ltd. All rights reserved.

***DOI 10.1016/S1367-5931(02)00393-9***

**Abbreviation**
**TM** transmembrane

## Introduction

Helical membrane proteins, and often their topology, can be reliably predicted by computation [1–3]. The challenge now is to understand how transmembrane (TM) helices fold to form 3-D structure. A critical step of membrane-protein folding is the assembly of TM helices. In this review, we examine progress made in the past few years about the nature of the driving force and the determinants of TM helical association.

The mechanism of the assembly of TM helices is complex. An early suggestion about membrane proteins is that they are 'inside-out' soluble proteins with polar cores and non-polar exteriors facing the lipid. However, structural analysis indicates that the distribution pattern of polar residues is more complex [4,5]. It was also thought that interacting helices are mostly nearest-neighbouring helices in primary sequences. This rule works perfectly for bacteriorhodopsin, but the structures of complex membrane proteins showed that it is not straightforward to predict which pair of helices interact. For example, calcium-transporting ATPase (pdb: 1eul) contains 10 helices, and has about 20 different helix–helix interactions. Among these, only seven are between connected helices [6]. Furthermore, not all connected helices interact: there are no interactions between helices 2 and 3 and helices 6 and 7 for this protein [7••].

## Packing density and voids in membrane proteins

On the basis of the structure of bacteriorhodpsin, Luecke *et al.* [8] suggested that membrane proteins are packed tighter than soluble proteins. Eilers *et al.* [9•,10] used the technique of occluded surfaces to show that the packing densities of residues in membrane proteins are generally higher than those of soluble proteins. These findings are consistent with the fact that lateral pressure is exerted to membrane proteins at the interface between protein and lipid bilayer. Gly, Pro and Ala have the highest packing values [10], as well as the highest coordination numbers [7••], indicating the important role of these small residues in facilitating helix–helix interactions

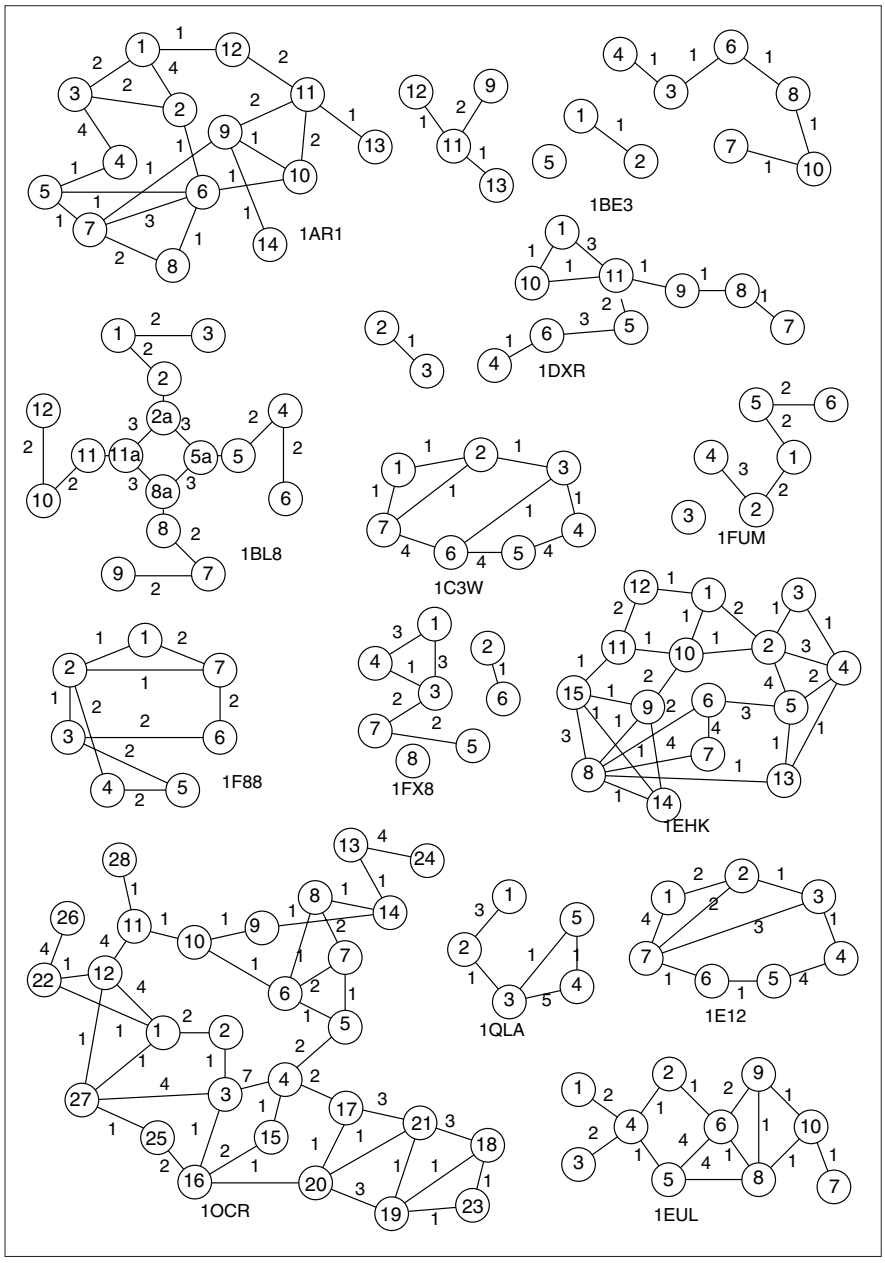
Although membrane proteins are packed tightly, they also contain numerous voids and pockets [7••]. These packing defects provide spaces for binding ligands, prosthetic groups, lipids and water molecules, and facilitate conformational changes essential for protein function. This is similar to soluble proteins, because many packing defects are found [11]. Small-to-large mutations in the central cavity of the channel of M2 protein from influenza A virus only causes a modest change in protein stability [12••]. These observations point to the hypothesis that proteins are not optimized by evolution to eliminate voids.

## Composition- and position-dependent single-residue propensity

Membrane proteins have simple amino acid residue composition. Genome-wide analysis of predicted TM helices reveals residue bias at different locations of the helices and in different organisms [13], confirming experimental data (see [14,15] for reviews). The canonical α-helical structures are often modified by Pro, Gly, pi-bulge and other structural elements [16,17], which may be important for protein stability and function. The role of Gly in facilitating helix–helix interactions, especially at helix crossing points, has been well studied [18,19]. Bending introduced by Pro may be necessary for ion-conducting [16]. A recent study of cystic fibrosis TM conductance regulator suggested that, alternatively, Pro residues may be selected during evolution to maximizes proper folding of nascent TM segments in the aqueous phase of the translocan against β-sheet aggregates before insertion into membrane [20•,21].

An interesting question is whether there is a strong bias for different residues to face lipid or to interact with other helices. By comparing the differential frequencies of amino acid residues in single-span and polytopic membrane proteins, Pilpel *et al.* [22] developed a scale for residues to face lipid environment. Aliphatic residues tend to face lipid, and the lipid propensity of residues is in general agreement with hydrophobicity scale. Residues not facing the lipid environment are more likely to interact with residues from another helix. With correction of unequal occurrence of residues in the TM helices, it was found recently that the small residues Ala, Gly and Ser have high propensity for helix–helix interactions [10], and Arg

**Figure 1**



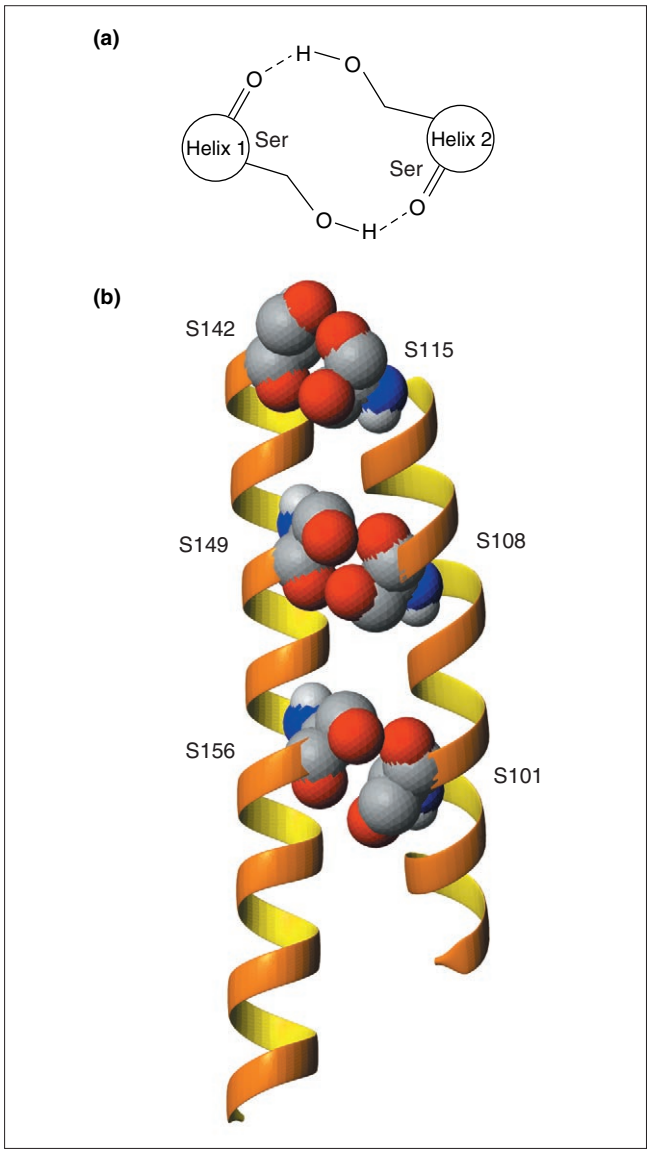

The topological maps of interhelical H-bonds (top view) of the TM region for 13 membrane proteins. Two helices are connected by an edge if there is at least one H-bond between them. The number of observed interhelical H-bonds is indicated next to the corresponding edge. The helices are numbered by the order in which they appear in the primary sequence. Reproduced from [34••], with permission. Labels underneath structures are Protein Data Bank codes. *Proteins* © 2002 John Wiley & Sons, Inc.

and Gln are less likely to be involved in interhelical interactions [7••].

## H-bonding in transmembrane regions

The role of van der Waals packing in maintaining membrane protein folding and assembly has long been recognized [23]. For example, a heptad motif of Leu residues can drive the association of designed membrane proteins with single TM helices [24]. The critical role of polar interactions in providing stability to helix association was recently highlighted in two important experiments using an engineered leucine zipper [25••,26••]. After replacing non-buried residues with hydrophobic residues (Ala, Val and Leu), engineered leucine zipper can be solubilized in micelles of SDS and non-ionic detergents, as well as in biological membranes, where they self-associate to form oligomers. However, the oligomerization is abolished once an Asn residue at position 14 is mutated to Val, as assayed by analytical ultracentrifugation, fluorescence, and circular dichroism spectroscopies. NMR studies showed that a strong hydrogen bond is formed between the two Asn residues at the dimer interface [26••]. Oligomerization also occurs when Asn is replaced by other residues with two polar side-chain atoms (Asp, Gln and Glu), but is undetectable when replaced with residues of one polar side chain atom (Ser and Thr) and hydrophobes in this system [27•]. These experiments suggest that hydrogen bonds can drive helical assembly. Further oligomerization experiments using helix of polyleucine showed that in addition to

**Figure 2**




Serine zipper as an example of spatial motif. **(a)** Schematic representation of H-bond between two serine residues on two adjacent helices. The side chain of each serine residue forms a H-bond with the carbonyl oxygen of its counterpart. Two side-chain-to-backbone H-bonds are formed. **(b)** Serine zipper in bovine cytochrome *c* oxidase between helices III and IV. Reproduced from [34••], with permission. *Proteins* © 2002 John Wiley & Sons, Inc.

Asp, Asn, Glu and Gln, His can also drive the assembly of TM helices [28].

Hydrogen bonds in a low dielectric environment have larger free energy gain than in aqueous solution [29]. Regular secondary structure elements (α-helices and β-sheet) can satisfy the main-chain hydrogen bonding potentials and therefore dominate in the lipid environment. For β-sheet, H-bond donor and acceptor groups can be saturated by H-bonding between adjacent β strands. For α-helices, it was originally thought H-bonds would predominantly exist intrahelically. The importance of interhelical polar interactions is now firmly established experimentally in model systems [27•,28], but to what extent is interhelical hydrogen bond found in natural helical membrane proteins? Polar and ionizable residues constitute *20–22%* of all residues in the TM helices [30,31••]. At least one hydrogen bond exists between each pair of adjacent helices in bacteriorhodopsin [8]. An H-bond network involving several polar residues was also found in the crystal structure of rhodopsin [32]. Polar groups are thought to be used by nature sparingly for interhelical interactions, since it might cause unintended strong interhelical H-bonds between helices with disastrous biological consequences [33]. In a recent computational analysis of a set of 13 structures of membrane proteins, however, it was found that there are extensive H-bond connections between helices: almost all TM helices are involved in interhelical H-bonding (Figure 1) [34••]. Among all 296 interacting helical pairs for these 13 proteins, one or more H-bond exists in 53% (158) of pairs. Side chains from Ser, Tyr, Thr and His are found in 50% of the H-bonds. In addition, the total interhelical atomic contacts are also different for helical pairs with and without H-bonds. On average, helical pairs with H-bonds are packed tighter and have more atomic contacts.

The importance of H-bonds in interhelical interactions is also confirmed in glycophorin A. The formation of glycophorin A dimer in membrane involves a well-packed interface. A recent study by Senes *et al.* [31••] suggests that, in addition to van der Waals interactions, weak $C_\alpha$-H–O H-bonds are important for the dimerization of glycophorin A. The analysis of a set of 11 membrane proteins showed that $C_\alpha$-H–O hydrogen bonds are very common, and are strongly favoured by right-handed parallel helical interactions. They are found to cluster around Gly, Ser and Thr residues. Polarized FT-IR and solid-state NMR studies indicated that, in addition to weak hydrogen bonds, the β-hydroxyl group of Thr87 also H-bonds with the backbone carbonyl group of Val84 [35].

Interhelical H-bonds are important for stability, assembly, and sometimes functions of polytopic membrane proteins. One attractive hypothesis is that hydrogen bonding provides the necessary driving force for association, which allows less extensive van der Waals packing interactions to provide the needed flexibility for movement important for function [28]. Residues with two side chain polar atoms (Asn, Asp, Gln and Glu) are rarely found in TM helices [27•]. However, polar residues Gln, Asn and His in TM regions experience purifying selection pressure and are highly conserved, suggesting their important role in maintaining membrane-protein structure integrity and/or biological function [36•].

## Pairwise interhelical interaction

The overall pattern of helix–helix interactions has been examined in several studies [7••,10,34••,37]. The contact plot technique was developed for studying the helix–helix interface [10,37]. This method calculates interatomic distances between backbone atoms of interacting helix pairs, and

identifies the helix–helix interface as residues occurring within 0.5 Å of a local minimum of the distances. An alternative method is to compute the alpha shape of membrane-protein structure and identify contacting residues as those sharing Voronoi planes [7••]. Studies using both techniques came to the same conclusion: membrane proteins have a far more diverse pattern of helical residue–residue interactions than soluble proteins [7••,10].

There are many differences between the patterns of helix–helix packing interactions for membrane and for soluble proteins. For example, Cys–Cys interactions are very common in soluble proteins, but are rarely seen in membrane proteins [7••,38]. A simple probabilistic model has been introduced to quantitatively assess membrane-helical interfacial pairwise contact propensity for specific residue pairs [7••]. Perhaps the most obvious pattern is that among residue pairs with high propensity for interhelical interactions; many are between polar or ionizable residues. Polar–polar interactions in membrane proteins are far more diverse than in soluble proteins [7••], a fact consistent with the prevalent and important roles H-bonds play in maintaining helical assembly in membrane proteins.

Much of the interhelical packing is due to side-chain–side-chain interactions. Gly–Gly is the only residue pair that shows a high propensity for backbone-atom interactions. The explicit consideration of side-chain atoms is important. For example, glycophorin A and engineered GCN4 leucine zipper represent two modes of helix association that are shared by other membrane proteins. The two main chains of GpA adopt a splayed configuration and the two chains of leucine zipper tightly wrap around each other forming a coiled-coil association. By the criterion of main-chain packing, the TM helices of leucine zipper pack more tightly than helices in GpA. However, a detailed examination of packing of the side chains using the alpha shape method showed that these two proteins have essentially the same packing efficiency. The average numbers of contacting atoms per atom in the interface are the same for these two proteins [7••].

It is intriguing to consider the possible roles of higher order interhelical interactions, which surely exist in membrane proteins. Because the scarcity of membrane protein structures, the technical challenge is to estimate accurately at least some of the 1540 possible parameters for three-body interactions. This situation may soon improve with the accumulation of more membrane protein structures, and perhaps with the help of statistical modeling, such as bootstrap resampling with various additional variance stabilization techniques [39,40].

## Sequence and structure motifs

To identify sequence motifs that mediate helix–helix interaction, Senes *et al.* [31••] examined exhaustively the frequency of all pairs and triplets of residues in the predicted sequences of individual TM helices, and identified a large number of sequence motifs that are significantly over-represented in membrane proteins. These include GG4 (GxxxG), GA4, and many other motifs utilizing small residues (Gly, Ala and Ser) at position *i* and *i+4*, often in association with a large residue at *i±1* and *i±2* positions. Among these, the GG4 motif is well known for mediating GpA dimerization. Screening tests using the TOXCAT system for parallel homodimerization with a large randomized peptide library showed that many peptides containing the GG4 motif have high affinity for helix–helix association [19]. Further experiments showed that oligomerization of homo helices can be facilitated by other sequence motifs such as SxxSSxxT and SxxxSSxxT [41]. Recent statisticaly analysis of helices in soluble proteins showed that GG4 is also a prevalent sequence motif stabilizing helix–helix association in soluble proteins [42].

Although the 'knobs-into-holes' spatial pattern exemplified by the leucine zipper is a well-studied general spatial pattern for helix interaction, the structural features underlying many sequence motifs described in [31••] are not well known. Senes *et al.* [43••] identified a connected network of $C_\alpha$-H–O hydrogen bonds, which are frequently found in right-handed parallel helix pairs enriched by Gly, Ser and Thr. One or two GG4 sequence motifs often participate in the network, and it is likely that the GG4 motif drives association of TM helices, in part by promoting $C_\alpha$-H–O=C interactions [43••]. The $C_\alpha$-H–O bond also stabilizes helix association in soluble proteins [42].

Serine residues have a high propensity to form self-pairs. It was thought that Ser mostly behaves as a nonpolar residue because of its tendency to adopt intrahelical hydrogen bonds with the backbone. However, it is possible that interhelical Ser–Ser interactions may serve as a basic unit of interaction behind some of the sequence motifs involving Ser residues. Experimental work has shown that the hydroxyl groups are important in the Sxx(x)SSxxT motifs, indicating the requirement of an extended H-bond network [41]. Although a single Ser–Ser contact may not provide adequate affinity for association, the cooperative interactions involving several Ser–Ser contacts may provide the needed stability and specificity. This is supported by the recent discovery of the spatial motif of the *serine zipper* [34••]. In a spatial arrangement analogous to the leucine zipper, a series of Ser–Ser H-bonds between two helices can form a serine zipper consisting of multiple Ser–Ser interactions with a periodicity of seven, which facilitates a tight association with five or more H-bonds [34••]. This serine zipper spatial motif is found in the structure of bovine cytochrome *c* oxidase. Sequence alignment using PSI-BLAST on a non-redundant database of TM domains [31••] suggests that serine zipper exists in other cytochrome *c* oxidase proteins whose structures are yet to be solved [34••]. It is likely that the sequence motif SS7 identified in [31••] may be directly related to the serine zipper spatial motif.

'Polar clamp' is another recently discovered spatial motif [34••]. It is a general motif that provides favourable

stabilization, and may enhance specific orientation of side chains in the background of generic nonpolar interactions. In this motif, a residue capable of forming two H-bonds (e.g. Glu, Lys, Asn, Gln) is clamped by two interhelical H-bonds formed with two other residues (e.g. Ser) located on a neighbouring helix at positions *i* and *i+1…i+4*. Polar clamps are found in all of the 13 proteins studied in [34••], and may be important for protein functions. In many cases, the SS4 sequence motif identified by Senes *et al.* is involved in the polar clamp spatial motif.

## Energy function and conformation sampling

Even with better understanding of helix–helix interactions in the membrane, the task of predicting the 3-D structure of membrane proteins remain daunting, especially for polytopic membrane proteins with many TM helices. Success will probably depend on two critical components: an efficient method that generates likely candidate conformations, and a potential function that effectively discriminates native structures from all other structures. It is important to keep in mind that wild-type membrane proteins are not necessarily optimized for stability [44•], and it is possible to engineer hyperstable membrane proteins with just a few mutations [45].

Potential functions based on various physical models have been applied in computational studies of membrane-protein folding in 2-D and 3-D lattice models [46•,47]. The physical interactions considered include hydrophobicity, bending, H-bonding [47], a modified HP model [48], and a modified Lennard–Jones Go-potential [46•]. These potentials are very informative in painting a likely scenario of the assembly of TM helices. For example, simulation of the folding of the first two TM helices of bacteriorhodopsin using Go-like potential indicates the existence of intermediate traps and provides details of multi-state kinetics. Nevertheless, the effectiveness of these potentials in discriminating native membrane-protein structures from other alternative structures is inadequate or unknown. Statistical parameters such as those developed from known protein structures provide an alternative approach for obtaining potential for predicting the coarse grained structure of membrane proteins. These parameters might include propensity for lipid exposure, pairwise propensity for interhelical interactions such as those developed in [7••], and possibly parameters for higher-order interactions in the future. Once more structures of membrane proteins become available, pairwise propensity and other parameters could be made position-dependent, the values of which will depend on the distance of their locations to the membrane–solution interface.

A third approach for developing effective potential is by optimization. This approach has been well developed in the studies of soluble-protein folding. The idea is simple: an energy or scoring function must have a lower value for the native structure than for any other alternative decoy structures. When the scoring function takes the empirical form of a linear sum of pairwise contacts with coefficients to be determined, the optimal coefficients can be obtained by optimization methods such as perception learning and linear programming [49,50]. State-of-the-art implementations of optimization methods such as the interior-point method for linear programming [51,52] are often helpful. A lesson learned in these studies relevant for membrane proteins is that parameters of higher-order interactions may be a necessity for successful folding of proteins [49,50].

For membrane-protein structure prediction, effective sampling of possible conformations is critical. These conformations provide the candidate pool to fish out the native-like structure. They can also provide challenging decoys for developing more discriminating scoring potentials via optimization. The task is to generate a large number of conformations that are protein-like. Despite the exponentially increasing number of ways to arrange TM helices, an important study showed that it is possible to generate computationally a template fold with up to seven canonical helices [53•].

The basic requirements for generating conformations is that they have to be self-avoiding walks in 3-D space, and they need to be membrane-protein like. The generation of a full atomic structure is challenging because of the large search space. It is often convenient to use the coarse grain model and start from a random or a modelled structure, and to use Markov chains with various move sets and Metropolis rules to generate an ensemble of conformations [47]. To obtain a set of samples following Boltzmann distribution of the chosen scoring function would require significant amounts of burn-in time before conformations can be harvested.

A promising alternative approach for generating conformation is the *chain growth method* [54,55]. This method grows a polymer chain *de novo* by adding monomers one at a time under the potential function. The advantage is that a larger region of conformational space can be searched, because conformations obtained are independent and uncorrelated. A challenge for chain growth method is the attrition problem: For soluble proteins, it is difficult to grow long self-avoiding chains. For membrane proteins, it is difficult to grow an assembly of self-avoiding chains that are packed tightly. Because of this limitation, this method has only been applied typically [54,55] to lattice models and coarse grained off-lattice models. Recent progress in various resampling strategies suggests that these limitations can be overcome. A more general approach is the Sequential Monte Carlo method, which allows importance sampling with arbitrary objective criteria and can achieve significantly improved efficiency in sampling [56]. These techniques will probably be useful for folding membrane proteins.

## Conclusions

Experimental and computational studies in the past few years have provided a much improved understanding

about the determinants of helix assembly of membrane proteins. It is clear now that polar residues and H-bonds play critical roles in the organization of membrane proteins. The geometry of packing in membrane proteins is also now well understood. A large number of sequence motifs and several spatial motifs promoting TM helical association are now known. We can expect that the energetics of helix association will be mapped out in more detail in the near future, for both model systems and membrane proteins with multiple TM helices. In addition, the search for higher-order interaction spatial patterns and their relationships with sequence motifs will probably bring additional insights into membrane-protein folding. With this progress, it is likely that experimental design of novel membrane proteins, computational studies of kinetics of folding, as well as efforts in solving the challenging problems of predicting 3-D structures of multispan membrane proteins on the basis of folding principles will intensify.

## Update

A recent Monte Carlo study showed that effective pairwise membrane potential can be obtained by optimization of native bR structure against decoy conformations generated iteratively. This potential can successfully fold bacteriorhodopsin. With additional modification, it can also fold glycophorin A [57].

## Acknowledgements

I thank Larisa Adamian for fruitful collaborations on studying membrane proteins. I am grateful to James Bowie, William DeGrado, Thomas Ebrey, Donald Engelman, Robert Gennis, Masahiro Kono, Gunnar von Heijne and Vikas Nanda for valuable discussions. This work is supported by funding from National Science Foundation CAREER DBI0133856, DBI0078270 and MCB998008, and the donors of the Petroleum Research Fund administered by the American Chemical Society (35616-G7).

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Rost B, Casadio R, Fariselli P, Sander C: **Transmembrane helices predicted at 95% accuracy.** *Protein Sci* 1995, **4**:521-533.

2. Rost B, Fariselli P, Casadio R: **Topology prediction for helical transmembrane proteins at 86% accuracy.** *Protein Sci* 1996, **5**:1704-1718.

3. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**:567-580.

4. Rees DC, Eisenberg D: **Turning a reference inside-out: commentary on an article by Stevens and Arkin entitled: "Are membrane proteins 'inside-out' proteins?" (*Proteins* 1999, 36:135-143).** *Proteins* 2000, **38**:121-122.

5. Stevens TJ, Arkin IT: **Turning an opinion inside-out: Rees and Eisenberg's commentary (*Proteins* 2000;38:121-122) on "Are membrane proteins 'inside-out' proteins?" (Proteins 1999;36:135-143).** *Proteins* 2000, **40**:463-464.

6. Toyoshima C, Nakasako M, Nomura OH: **Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution.** *Nature* 2000, **405**:647.

7. •• Adamian L, Liang J: **Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins.** *J Mol Biol* 2001, **311**:891-907.
This work surveys the packing geometry of TM helices in membrane proteins, including the distribution of voids and pockets, packing efficiencies, residue coordination numbers, and helix–helix interactions. It was found that there are extensive specific pairwise interhelical interactions, including aromatic-basic and polar–polar interactions. In addition, a membrane helical interfacial pairwise propensity scale was developed.

8. Luecke H, Schobert B, Richter H-T, Cartailler J-P, Lanyi JK: **Structure of bacteriorhodopsin at 1.55 Å resolution.** *J Mol Biol* 1999, **291**:899-911.

9. • Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ: **Internal packing of helical membrane proteins.** *Proc Natl Acad Sci USA* 2000, **97**:5796-5801.
Using the technique of occluded surface, membrane proteins are found to have higher packing values than soluble proteins.

10. Eilers M, Patel AB, Liu W, Smith SO: **Comparison of helix interactions in membrane and soluble alpha-bundle proteins.** *Biophys J* 2002, **82**:2720-2736.

11. Liang J, Dill KA: **Are proteins well-packed?** *Biophys J* 2001, **81**:751-766.

12. •• Howard KP, Lear JD, DeGrado WF: **Sequence determinants of the energetics of folding of a transmembrane four-helix-bundle protein.** *Proc Natl Acad Sci USA* 2002, **99**:8568-8572.
The first quantitative measurements of thermodynamics of sequence variants of TM helical bundle domains with a large number of helices. The effects of single mutations at core positions of the homotetramer TM domain of the M2 protein from influenza A virus were measured by analytical ultracentrifugation. Many of the mutations are found to increase the stability of helical association, suggesting that natural membrane proteins are not optimized by evolution for maximum stability for oligomerization.

13. Wallin E, von Heijne G: **Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms.** *Protein Sci* 1998, **7**:1029-1038.

14. White SH, Wimley WC: **Membrane protein folding and stability: physical principles.** *Annu Rev Biophys Biomol Struct* 1999, **28**:319-365.

15. Killian JA, von Heijne G: **How proteins adapt to a membrane-water interface.** *Trends Biochem Sci* 2000, **25**:429-434.

16. Tieleman DP, Shrivastava IH, Ulmschneider MR, Sansom MS: **Proline-induced hinges in transmembrane helices: possible roles in ion channel gating.** *Proteins* 2001, **44**:63-72.

17. Riek RP, Rigoutsos I, Novotny J, Graham RM: **Non-alpha-helical elements modulate polytopic membrane protein architecture.** *J Mol Biol* 2001, **306**:349-362.

18. Javadpour MM, Eilers M, Groesbeek M, Smith SO: **Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association.** *Biophysical J* 1999, **77**:1609-1618.

19. Russ WP, Engelman DM: **The GxxxG motif: a framework for transmembrane helix-helix association.** *J Mol Biol* 2000, **296**:911-919.

20. • Wigley WC, Corboy MJ, Cutler TD, Thibodeau PH, Oldan J, Lee MG, Rizo J, Hunt JF, Thomas PJ: **A protein sequence that can encode native structure by disfavoring alternate conformations.** *Nat Struct Biol* 2002, **9**:381-388.
A Pro residue in the centre of the third TM helix of the cystic fibrosis conductance regulator is found to promote folding by disfavouring misfolded β-sheet aggregate. On the basis of genome-wide TM sequence analysis, the authors suggest that this is a general mechanism selected by membrane proteins.

21. Deber CM, Therien AG: **Putting the beta-breaks on membrane protein misfolding.** *Nat Struct Biol* 2002, **9**:318-319.

22. Pilpel Y, Ben-Tal N, Lancet D: **kPROT: a knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein structure prediction.** *J Mol Biol* 1999, **294**:921-935.

23. MacKenzie KR, Prestegard JH, Engelman DM: **A transmembrane helix dimer: structure and implications.** *Science* 1997, **276**:131-133.

24. Gurezka R, Laage R, Brosig B, Langosch D: **A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments.** *J Biol Chem* 1999, **274**:9265-9270.

25. •• Choma C, Gratkowski H, Lear JD, DeGrado WF: **Asparagine mediated self-association of a model transmembrane helix.** *Nat Struct Biol* 2000, **7**:161-166.
One of the first studies elucidating the important role of polar–polar interactions in TM helix assembly. A two-stranded coiled-coil peptide based on

yeast transcription factor GCN4 leucine zipper was engineered to be membrane-soluble. The oligomerization of this peptide was found to critically depend on the presence of a single Asn residue. Mutation to Val essentially abolishes oligomerization.

26. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM:
•• **Interhelical hydrogen bonding drives strong interactions in membrane proteins.** *Nat Struct Biol* 2000, **7**:154-160.
One of the first studies demonstrating the important role of polar residues in the association of TM helices. Using engineered leucine zipper, the Asn residue was found to be essential for helix association in both detergent and biological membrane.

27. Gratkowski H, Lear JD, DeGrado WF: **Polar side chains drive the**
• **association of model transmembrane peptides.** *Proc Natl Acad Sci USA* 2001, **98**:880-885.
Building on oligomerization experiments of a model TM helix of engineered GCN4-P1, this work showed that residues with two polar atoms can drive oligomerization, whereas Ser and Thr cannot. The authors also discuss these results in light of the frequencies of polar residues appearing in TM helices.

28. Zhou FX, Merianos HJ, Brunger AT, Engelman DM: **Polar residues drive association of polyleucine transmembrane helices.** *Proc Natl Acad Sci USA* 2001, **98**:2250-2255.

29. Pace CN: **Polar group burial contributes more to protein stability than nonpolar group burial.** *Biochemistry* 2001, **40**:310-313.

30. Arkin IT, Brunger AT: **Statistical analysis of predicted transmembrane α-helices.** *Biochim Biophy Acta* 1998, **1429**:113-128.

31. Senes A, Gerstein M, Engelman DM: **Statistical analysis of amino**
•• **acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions.** *J Mol Biol* 2000, **296**:921-936.
Residue pairs and triplets with different sequence separation in individual TM helices were exhaustively examined for a large set of predicted TM helices. Many sequence motifs that are significantly over-represented were found, including GG4, SS4 and many others involving small residues.

32. Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Trong I, Teller DC, Okada T, Stenkamp RE *et al.*: **Crystal structure of rhodopsin: a G protein-coupled receptor.** *Science* 2000, **289**:739-745.

33. Smith SO, Smith CS, Bormann B: **Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbB-2 receptor.** *Nat Struct Biol* 1996, **3**:252-258.

34. Adamian L, Liang J: **Interhelical hydrogen bonds and spatial motifs**
•• **in membrane proteins: polar clamps and serine zippers.** *Proteins* 2002, **47**:209-218.
A comprehensive survey of interhelical H-bond interactions shows that there are extensive H-bond connections between interacting TM helices. Almost all TM helices have one or more interhelical H-bond, and about 50% of interacting helical pairs contain H-bonds. Interhelical interfaces with H-bonds, in general, are packed tighter than those without H-bonds. Several spatial interhelical motifs, including serine zipper and polar clamps, are discovered.

35. Smith SO, Eilers M, Song D, Crocker E, Ying W, Groesbeek M, Metz G, Ziliox M, Aimoto S: **Implications of threonine hydrogen bonding in the glycophorin a transmembrane helix dimer.** *Biophys J* 2002, **82**:2476-2486.

36. Tourasse NJ, Li WH: **Selective constraints, amino acid**
• **composition, and the rate of protein evolution.** *Mol Biol Evol* 2000, **17**:656-664.
A careful comparative study of substitution rates of amino acid residues in the TM region and in soluble proteins. Using a set of homologous membrane proteins that have not yet experienced extensive multiple mutation, the authors found that polar residues Gln, Asn and His in TM regions experience purifying selection pressure and are highly conserved.

37. Javadpour MM, Eilers M, Groesbeek M, Smith SO: **Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association.** *Biophys J* 1999, **77**:1609-1618.

38. Popot JL, Engelman DM: **Helical membrane protein folding, stability, and evolution.** *Annu Rev Biochem* 2000, **69**:881-922.

39. Efron B, Tibshirani RJ: *An Introduction to the Bootstrap.* Boca Raton, Florida: Chapman and Hall/CRC; 1994.

40. Davison AC, Hinkley DV: *Bootstrap Methods and Their Application*: Cambridge, UK: Cambridge Univ Press; 1997.

41. Dawson JP, Weinger JS, Engelman DM: **Motifs of serine and threonine can drive association of transmembrane helices.** *J Mol Biol* 2002, **316**:799-805.

42. Kleiger G, Grothe R, Mallick P, Eisenberg D: **GXXXG and AXXXA: common α-helical interaction motifs in proteins, particularly in extremophiles.** *Biochemistry* 2002, **41**:5990-5997.

43. Senes A, Ubarretxena-Belandia I, Engelman DM: **The $C_\alpha$–H.O**
•• **hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions.** *Proc Natl Acad Sci USA* 2001, **98**:9056-9061.
This paper examined systematically $C_\alpha$-H–O hydrogen bonds in a set of 11 membrane proteins. Extensive networks of interactions involving $C_\alpha$-H–O bonds are found in many proteins. These often involve one or two GG4 sequence motifs and small residues such as Gly, Ser and Thr. A strong preference for parallel right-handed helices is found, and it was suggested that several coordinated $C_\alpha$-H–O bond interactions are important for the stability and specificity of helix–helix association in membrane proteins.

44. Bowie JU: **Stabilizing membrane proteins.** *Curr Opin Struct Biol*
• 2001, **11**:397-402.
An insightful review on how stability-enhancing mutations are relatively easy to find. The author suggests that membrane proteins are not optimized for stability, and points to experimental demonstration that a hyperstable membrane protein can be obtained by a quadruple mutation.

45. Zhou Y, Bowie JU: **Building a thermostable membrane protein.** *J Biol Chem* 2000, **275**:6975-6979.

46. Orlandini E, Seno F, Banavar JR, Laio A, Maritan A: **Deciphering the**
• **folding kinetics of transmembrane helical proteins.** *Proc Natl Acad Sci USA* 2000, **97**:14229-14234.
Using $C_\alpha$ representation of the first 66 amino acid residues from bacteriorhodopsin, which are known to fold independently, a Monte Carlo simulation was carried out to assess the folding kinetics and equilibrium thermodynamics of these two TM helices. Under Go-like potential encoding information about native contacts, the results of simulation suggest the existence of intermediates and multi-state folding kinetics.

47. Chen C-M: **Lattice model of transmembrane polypeptide folding.** *Phys Rev E* 2001, **63**:01090.1-4.

48. Bonaccini R, Seno F: **Simple model to study insertion of a protein into a membrane.** *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 1999, **60**:7290-7298.

49. Vendruscolo M, Najmanovich R, Domany E: **Can a pairwise contact potential stabilize native protein folds against decoys obtained by threading?** *Proteins* 2000, **38**:134-148.

50. Tobi D, Shafran G, Linial N, Elber R: **On the design and analysis of protein folding potentials.** *Proteins* 2000, **40**:71-85.

51. Meller J, Wagner M, Elber R: **Maximum feasibility guideline in the design and analysis of protein folding potentials.** *J Comput Chem* 2002, **23**:111-118.

52. Meller J, Elber R: **Linear programming optimization and a double statistical filter for protein threading protocols.** *Proteins* 2001, **45**:241-261.

53. Bowie JU: **Helix-bundle membrane protein fold templates.** *Protein*
• *Sci* 1999, **8**:2711-2719.
This work explored the size of the space for conformational search in packing canonical helices into bundles. Fold template can be generated computationally for up to seven helices. The number of possible fold templates is found to increase exponentially with the number of helices.

54. Meirovitch H: **Scanning method as an unbiased simulation technique and its application to the study of self-attracting random walks.** *Physical Rev A* 1985, **32**:3699-3708.

55. Gan HH, Tropsha A, Schlick T: **Lattice protein folding with two and four-body statistical potentials.** *Proteins* 2001, **43**:161-174.

56. Liang J, Zhang J, Chen R: **Statistical geometry of packing defects of lattice chain polymer from enumeration and sequential Monte Carlo method.** *J Chem Phys* 2002, **117**:3511-3521.

57. Dobbs H, Orlandini E, Bonaccini R, Seno F: **Optimal potentials for predicting inter-helical packing in transmembrane proteins.** *Proteins* 2002, **49**:342-349.