# Natural constraints, folding, motion, and structural stability in transmembrane helical proteins

**ARTICLE** · JANUARY 2010

DOI: 10.1007/978-3-7091-0045-5_12



**2 AUTHORS**, INCLUDING:

Nir Ben-Tal

Tel Aviv University

**206** PUBLICATIONS **8,119** CITATIONS

# Natural constraints, folding, motion, and structural stability in transmembrane helical proteins

**Susan E. Harrington and Nir Ben-Tal**

Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Israel

**Abstract**

Transmembrane (TM) helical proteins are of fundamental importance in many diverse biological processes. To understand these proteins functionally, it is necessary to characterize the forces that stabilize them. What are these forces (both within the protein itself and between the protein and membrane) and how do they give rise to the multiple conformational states and complex activity of TM helical proteins? How do they act in concert to fold TM helical proteins, create their low-energy stable states, and guide their motion? These central questions have led to the description of critical natural constraints and partial answers, which we will review. We will then describe how these constraints can be tracked through homologs and proteins of similar folds in order to better understand how amino acid sequence can specify structure and guide motion. Our emphasis throughout will be on structural features of TM helix bundles themselves, but we will also sketch the membrane-related aspects of these questions.

## 1 Folding background

Central aspects of transmembrane (TM) helical protein folding are well understood and have important structural implications (Bowie 2005).

### 1.1 Two-stage hypothesis

The widely accepted two-stage hypothesis provides the foundation for much current work in TM helical protein folding and structure prediction. It states that TM

*Corresponding author*: Susan E. Harrington, Department of Biochemistry and Molecular Biology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel (E-mail: susan.e.harrington@gmail.com)

helices fold autonomously in the first stage of folding within the membrane, then associate to form helix pairs, then triples or quadruples and so forth, eventually building up to the full bundle (Popot and Engelman 1990). The two-stage hypothesis is known (and was known by its creators) to be a simplification, but it is a very useful one. It has led to the fruitful investigation of the association of helix pairs as the foundation for the folding of the full bundle and the structure prediction of TM helical proteins.

### 1.2 Translocon-aided folding

As they come off the ribosome during translation, TM helical proteins enter the membrane (the endoplasmic reticulum membrane in eukaryotes or the plasma membrane of bacteria) via the translocon complex, and the translocon inserts the TM helices into the membrane in sequence order (usually; Sadlish et al. 2005; White and von Heijne 2008). Experimental evidence suggests that the translocon can measure the hydrophobicities of helices to determine which are hydrophobic enough to enter the membrane rather than be secreted, and the associated natural translocon hydrophobicity scale has been derived (Hessa et al. 2005, 2007). Interestingly, according to this scale, a significant fraction of TM helices are insufficiently hydrophobic to be inserted by the translocon as isolated helices but must be inserted in sequence context with their immediately neighboring loops and helices (Hedin et al. 2010). On occasion this is also insufficient, and a more complex interplay between the protein and the translocon must be at work (Hedin et al. 2010). For the folding of some proteins, there is evidence for a much more elaborate and active role for the translocon–ribosome complex than is postulated in passive sequential-insertion models (Kida et al. 2007; Skach 2007; Pitonzo et al. 2009). Clearly, understanding the full function and mechanisms of the translocon will be an important challenge for years to come. Our chief interest here will be the influence of the translocon on the likely contact order of the TM helices and some of the possible structural consequences for the full helix bundles.

## 2 Overview of non-interhelical stabilizing forces and natural constraints

We first sketch some important non-interhelical interactions. These constraints and interactions will act cooperatively with the interhelical ones that will be our focus.

### 2.1 Membrane constraints and interactions

The membrane creates some of the foremost natural constraints for TM helical proteins. First, the membrane greatly limits the amino acid composition of a TM helical

protein: the protein must be sufficiently hydrophobic to insert into the membrane, yet it must also accommodate the membrane's polar headgroup region. Second, we have the following known geometric constraints.

Beyond these types of protein–membrane constraints, some TM helical proteins have evolved to respond to subtle changes in the membrane including lateral pressure, curvature, lipid composition, and phase, etc. (Perozo et al. 2002; Lundbaek et al. 2010).

#### 2.1.1 Hydrophobic mismatch

A TM helical protein must avoid hydrophobic mismatch with the membrane: the hydrophobic stretch of each TM helix will usually position itself to match the hydrophobic thickness of the membrane. Thus once in the membrane, a helix will usually have a restricted tilt angle with respect to the membrane normal. The specifics are being studied both experimentally and computationally (Bond et al. 2007; Krishnakumar and London 2007; Holt and Killian 2010).

#### 2.1.2 Specific flanking and anchoring interactions with polar headgroups

Residues with a mixed polar/apolar character are common in interfacial regions, and both basic and aromatic residues are reported to form specific favorable interactions with lipid headgroups (Ren et al. 1999; Killian and von Heijne 2000; Strandberg et al. 2002; Chamberlain et al. 2004). Trp anchoring is perhaps best established and has been studied experimentally in synthetic peptides in bilayers, where it is reported to inhibit helix tilting (de Planque et al. 2003; Chiang et al. 2005). Better characterization of protein–membrane interaction motifs would be an important advance.

#### 2.1.3 Positive-inside rule

The positive-inside rule states that the cytoplasmic loops of TM helical proteins tend to be enriched in positively charged residues, and so this charge distribution determines the orientation of most TM helical proteins within the membrane (von Heijne 1989). The physical basis for this tendency is not well-understood and is currently being studied (van Klompenburg et al. 1997; Bogdanov et al. 2008).

### 2.2 Loop constraints

Loops are clearly important to translocon-aided folding and known to be essential in some cases (White and von Heijne 2008; Hedin et al. 2010).

The effect of loops on stability and folding has been studied in depth experimentally for bacteriorhodopsin (bR) and rhodopsin. In the bR studies, loops were systematically clipped and the resulting fragments were observed to reconstitute

a native-like structure, but with reduced stability (Huang et al. 1981; Liao et al. 1984; Popot et al. 1987; Marti 1998). Similar studies were conducted for rhodopsin that showed most of its loops could not be clipped without disrupting the folding of the protein (Albert and Litman 1978; Litman 1979; Ridge et al. 1995; Landin et al. 2001).

Many loops are short and stretched, and so impose a significant geometric constraint on the connected helices (Enosh et al. 2004; Tastan et al. 2009).

## 3 Interhelical interactions and constraints

We can now begin to discuss how iterative folding and amino acid sequence come together to specify stable structures via well-known specific interhelical interactions. The interhelical constraints act cooperatively with the ones discussed above.

### 3.1 Helix–helix packing

Crick first described how a simplified view of the surface features of an alpha-helix could be seen to restrict the crossing angle of a close-packed pair of helices (Crick 1953). If the side chains are considered as simple knobs and the spaces between the knobs as holes, then if a pair of helices is to be brought into extended close contact, a series of the knobs must fit into corresponding hole regions. From the geometry of alpha helices and simplifying assumptions, he derived crossing angles for both parallel and antiparallel helices that would enable this type of packing. While that was a groundbreaking analysis, to fully describe helix–helix packing requires a more nuanced approach, as one can see from the more varied helix packing in solved structures.

Walters and DeGrado (2006) have made a very thorough analysis of helix–helix packing in solved TM helical structures. They selected 445 helix pairs from 32 high-resolution protein structures, aligned each possible pair, and clustered them so that each member of a cluster was within 1.5 Å Cα rmsd of a reference centroid pair (the rmsds were computed over 10–14 residue stretches of the TM helices). They found that 29% of pairs fell into the most populous cluster, 74% of pairs fell into the top five most populous clusters, and 90% fell within the top 14. They also found significant amino acid propensities for specific positions in some helix pair clusters. We will reexamine the clusters from a slightly different perspective in a later section.

### 3.2 Motifs and stabilizing specific interactions

Statistical studies of sequences, examination of solved structures, theoretical analysis, and various types of experiments have led to the description of a small number of

common significantly stabilizing interhelical interaction types. It has become common practice for scientists to spotlight where these kinds of interactions occur when introducing new experimental structures or models of TM helical proteins. These are:

#### 3.2.1 Packing motifs

These are usually composed of close knob-in-hole type packed residues each with a side chain of limited conformational flexibility. Thus each such residue can fill a cavity and make van der Waals (VDW) and sometimes polar contacts without significant entropic losses. The most common of these include the famous GxxxG, leucine zippers, and variants of both (Gurezka et al. 1999; Russ and Engelman 2000; Senes et al. 2001; Schneider and Engelman 2004). There are other less common ones, e.g., using proline packing (Senes et al. 2004).

#### 3.2.2 Hydrogen bonds

Although interhelical hydrogen bonds appear to be weaker on average than was once thought, their strength varies greatly according to environment, and they can be significantly stabilizing (Zhou et al. 2000; Gratkowski et al. 2001; Arbely and Arkin 2004; Joh et al. 2008). Residues participating in such hydrogen bonds are often highly conserved (Hildebrand et al. 2008).

#### 3.2.3 Aromatic interactions

The edges of aromatic rings can be considered as weak donors and acceptors. Resulting aromatic interactions include the well-established and important cation–pi interactions, aromatic stacking, edge-to-face aromatic–aromatic interactions, and interactions with polar atoms where the edge acts as donor and the center as acceptor (Dougherty 1996; Johnson et al. 2007; Sal-Man et al. 2007; Nanda and Schmiedekamp 2008).

#### 3.2.4 Salt bridges

Rare in TM helical proteins, but can be significantly stabilizing (Honig and Hubbell 1984).

### 3.3 The five types of specific stabilizing interhelical interactions considered

We will focus on five types of stabilizing interhelical interactions when analyzing structures. All of the interactions lie within or very close to the inferred hydrocarbon region. Three are polar: hydrogen bonds, salt bridges, and some aromatic

interactions. Two are packing interactions: small residue (G/A/S/C) as knob in close knob-in-hole packing and I/V/L/T as knob in close knob-in-hole packing in I/V/L/T contact patches. For I/V/L/T patch packing, at least one of the surrounding hole residues must be I/V/L/T, along with some other restrictions on the hole residue types (Harrington and Ben-Tal 2009). Note that these packing interactions have been defined on a residue basis, so common packing motifs would consist of multiple such interactions.

For each interaction type, there is a fixed interaction geometry. These interaction geometries are fixed sets of geometric conditions that must be satisfied if the donor and acceptor are interacting (e.g., the usual conditions for a hydrogen bond; for the two packing interactions, the knob is considered the donor and the hole the acceptor; Harrington and Ben-Tal 2009).

### 3.4 Structural hot spots

An idea related to the above list of interactions is that of what is sometimes called a structural hot spot: a residue or residues making particularly favorable contributions to stability (Bogan and Thorn 1998; Fleming and Engelman 2001). We will call a residue "structurally hot" or "particularly stabilizing" when it makes interhelical contacts so that its contribution to the stability of the conformation is especially favorable, and much more favorable than would be expected for a typical residue in a typical conformation of the protein. It has been suggested that residues likely to be structurally hot make numerous favorable contacts: some VDW, but also other more specific favorable interactions (Gao and Li 2009). Such residues are sometimes termed "hub residues" (Pabuwal and Li 2008, 2009); e.g., a polar residue that makes numerous VDW contacts while also hydrogen bonding. This characterization of structurally hot residues has experimental support, as will be described in the next section, and indicates why some of the weaker specific interactions included in our list can still play a critical role in creating structural hot spots. This is clear if the weak interactions are likely stronger than generic VDW contacts would be for the atoms in the interactions.

The structural hot spot concept relates to the motif-type analysis of structures: one can try to describe a set of geometric conditions for a set of residues to satisfy (often specifying residue types of interacting partners) so that the resulting contribution of the residues in the motif is very likely significantly favorable to the stability of the structure. For the common motifs, it will usually be true that at least one residue of the motif will contribute significantly more to stability than is expected for a typical residue. To facilitate structural analysis, it is convenient to take the various packing motifs apart into the packing of their residues.

Theoretically, one would like a list of simple structural conditions to place on amino acids so that the probability that they make a significantly favorable contribution to stability is much higher than that of a typical residue making typical contacts. In practice, consideration of the types of favorable interactions described above and characterizations of common motifs are extremely useful approximations to an ideal set of descriptors. They are particularly useful because they do not depend on complex tertiary contacts. Two things should be kept in mind when using these simplified conditions. First, the energies of the interactions can vary greatly, and in the case of weak hydrogen bonds or other polar interactions, the bond by itself does not make a participating residue "structurally hot". But the existence of even a weak bond increases the probability that a participating residue makes a significant contribution to the stability of the conformation. Second, even optimally defined conditions would have a probabilistic implication of stability, and a residue satisfying them might not be stabilizing.

In reality, not all structurally hot residues are involved in interactions of one of the five types. Conspicuously absent are residues that simply make many VDW contacts. These will primarily be residues with many tertiary contacts, and such structural hot spots could be recognized only when the protein is assembled to a sufficient extent. This limits their use during most steps of iterative translocon-aided folding or structure prediction assembly.

### 3.5 Experimental data on residue contributions to stabilization

Faham et al. (2004) investigated the contributions of residues to stability in bR by systematically mutating the residues of helix B to alanine and measuring the thermodynamic stability of the mutants using an unfolding assay. We will interpret their data in terms of the five types of stabilizing interactions. (Faham et al. did not present their results in these terms but instead emphasized VDW interactions in their interpretation of the data.)

Each mutant was classified as either severely destabilized, moderately destabilized, minimally altered, or stabilized. Of the 24 residues mutated, 17 could be said to have some kind of interhelical contact (of greatly varying extents) with one or more of the other bR helices, and we will restrict our attention to those. (The structure used for our analysis is 1C3W.) The notation used below gives the interactions in donor–acceptor form.

Of the 17, four were severely destabilizing: F42A, I45A, T46A, and Y57A. All of these residues make particularly stabilizing interhelical interactions of the five types, and two of them make at least two, in line with the "hub residue" type of structural hot spot. The mutations do not create favorable interactions of the five types, but do

destroy the native ones. (F42: 42-CD1-96-OD1, 42-CE1-96-O; I45 I/V/L/T patch packing; T46:46-OG1-96-OD2; Y57: 57-CD2-13-O, 57-OH-212-OD2, 57-CE1-212-OD2.)

Five of the 17 mutants were moderately destabilizing: Y43A, T47A, I52A, F54A, and M60A. The first four of these also make at least one particularly stabilizing interhelical interaction of the five types. Again, the mutations do not create favorable interactions. (Y43:30-NZ-43-OH, 43-CD1-27-O (borderline); T47: I/V/L/T patch packing, 47-OG1-27-O (water bridged); I52: I/V/L/T patch packing (borderline); F54: 54-CD1-17-O.)

The exception is M60, which by its residue type cannot participate in an interaction of one of the five types. Interestingly, it makes extensive VDW contact with Y57 in the native structure. Y57A is the most destabilizing mutation of all, and Y57 does make three particularly stabilizing interactions as was described. It appears that the side chain of M60 guides Y57 to make those interactions, thereby further stabilizing the native position of Y57.

The five minimally altered mutants of the 17 were: L48A, P50A, T55A, L58A, and S59A. With the exception of P50, these residues make few contacts of any kind. While we did not include proline packing in our list of favorable interactions, it would have been fair to do so. Since proline close packing would be considered favorable because it has a small side chain of restricted mobility, one would not expect P50A to much change stability; it is a moderate substitution. None of the other mutations would either add or remove a favorable interaction of the five types.

The three stabilized mutants of the 17 were: V49A, M56A, and L61A. Both V49 and L61 act as close-packed knobs in favorable I/V/L/T patch interactions. The substitution changes both interactions to a small residue knob-in-hole interaction, and so V49A and L61A amount to a substitution of one kind of favorable packing interaction for another. That alanine packing is superior is consistent with results on zipper motifs in TM helix pairs suggesting that small residue close knob-in-hole packing is more stabilizing than I/V patch packing (Zhang et al. 2009). M56 does not participate in any of the interactions of the five types, but M56A will convert it to a small residue in close knob-in-hole packing. Thus this substitution adds a particularly favorable packing interaction.

Overall, this data and above analysis support the stabilizing significance of the interactions of the five types.

### 3.6 Particularly stabilizing interactions as geometric constraints

By examining solved structures for these kinds of particularly favorable interhelical interactions, one can see that they seem to be distributed in a surprisingly meaning-

ful way: the interactions appear to tightly constrain the packing of the helices in most solved TM helical proteins. To put this another way, it appears that if we significantly perturb the helix positions, we necessarily break some interactions in the set. If true, one might say a complete set of determining constraints/interactions of the five types has evolved to nearly fix each native state backbone conformation. We call sets of the five types of interhelical interactions "determining sets" when they nearly

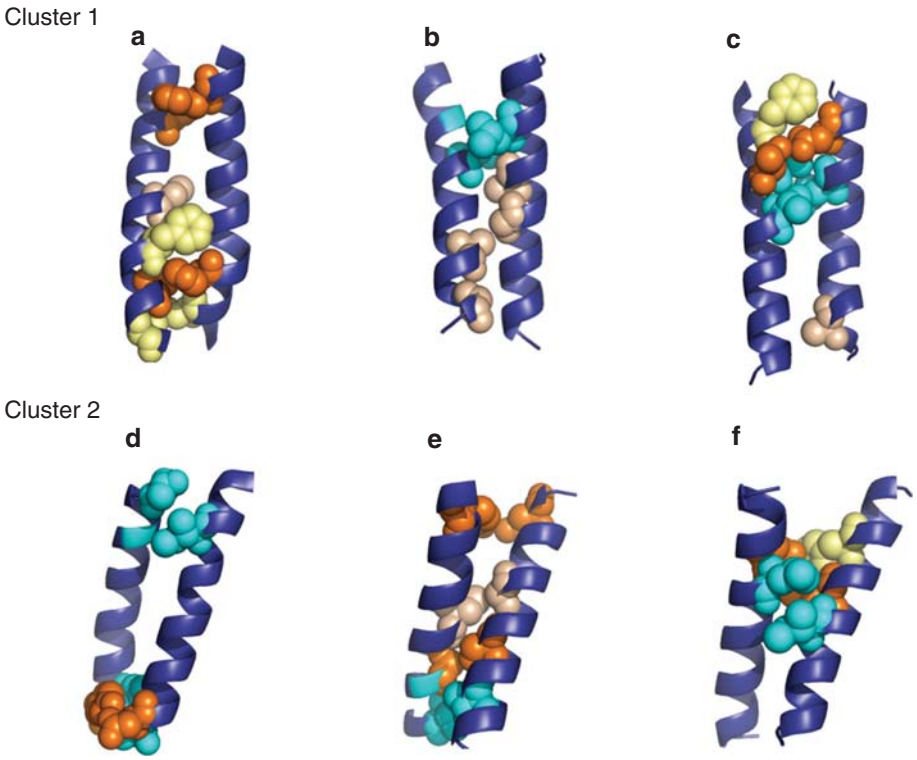


**Fig. 1. Unrelated members of the most common TM helix pair folds stabilized by diverse determining sets.** The five types of interhelical interactions are displayed as follows. (Only four types occur in these structures: there are no salt bridges.) Residues in a hydrogen bond are colored orange; if a residue's side chain (rather than a backbone atom) is in a hydrogen bond its atoms are shown as orange spheres. Residues in aromatic interactions are colored yellow; if their side chains participate in the interaction, the side chain atoms are shown as spheres. The knob atoms of G/A/S/C small close knob-in-hole packing are shown as wheat colored spheres. The knob atoms of I/V/L/T close knob-in-hole packing in I/V/L/T patches are shown as cyan (bright blue) spheres. The corresponding hole residues are shown as spheres if they are I/V/L/T in close contact with the knob residue; otherwise if the hole residue has restricted side chain conformations it is also shown in cyan. We can see unrelated helix pairs from the same fold with diverse determining sets. Upper panels: pairs from the top Walters–DeGrado cluster (**a**) 1Q90: cytochrome B6-F. (**b**) 1H2S: sensory rhodopsin II. (**c**) 1C3W: bR (a different helix pair than the homologous one for 1H2S). Lower panels: pairs from the next most populous cluster (**d**) 1OCR: cytochrome C oxidase. (**e**) 1RH5: translocase SecE subunit. (**f**) 1OKC: ADP/ATP carrier protein.

fix geometrically the packing of the helix backbones. (Our terminology is based on the geometric meaning of "determine": to specify position, to fix.)

We will discuss how these observations relate to some fundamental questions of protein folding and describe our work to establish them. But we first look at some simple examples.

### 3.7 Helix pairs revisited

The Walters–DeGrado clusters provide examples of diverse ways in which the five types of interactions can be distributed to fix the same helix-pair fold. The helix pairs within a cluster usually come from unrelated proteins or from different parts of the same protein. As members of the same cluster, they must have similar conformations for the aligned 10–14 residue region. Since most TM helices are at least 20 residues long, the position of this region can vary, and so for some clusters not all the full helix pairs look similar.

We have chosen pairs that do look similar and stable to show the diverse ways that a fold can be stabilized. In Fig. 1, we see three members from each of the two most populous clusters with their determining sets displayed. Imagine perturbing the positions of the helices to see how little one can move the helices without disrupting these interactions.

### 3.8 Constraint perspective and underlying rigid-body geometry

Both helix packing and the constraining effect of the five types of interactions can be better understood if some facts about the geometry of rigid bodies are kept in mind.

Any rigid body's position in space can be specified by the positions of any three non-collinear points on the body. If the exact positions of those three points are unknown, but we do know that they must each lie within three given regions in space, then we can obtain an initial ensemble of positions of the body by placing grids on those three regions and systematically selecting these points to give the positions of the three points on the body. These three positions then fix the position of the body itself. If there are additional restrictions on the positions in space of any other points on the body, then we can check the initial ensemble of positions of the body and remove any positions from the ensemble that do not meet those restrictions. By choosing sufficiently fine grids, one can find to any desired accuracy how the specified regions constrain the position of the body.

This approach can be adapted to build the combined piece ensemble of two pieces constrained by a set of any type of interactions with fixed known interaction regions. Three of these interactions and their associated regions can be used to position one piece relative to the other and build the initial combined piece ensemble. If there are

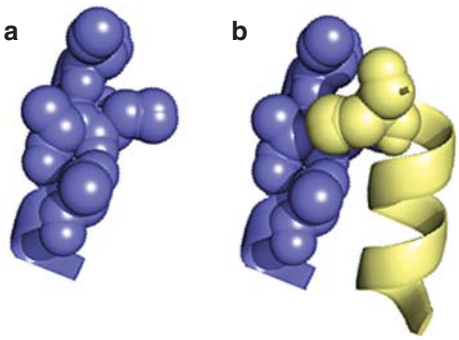


**Fig. 2.** **Knob-in-hole.** (**a**) The residues $j$, $j+3$, $j+4$, and $j+7$ on an alpha helix with reduced representations of their side chains with their atoms shown as spheres. The numbering starts from the top in this picture, and the space surrounded by these residues is called a hole. (**b**) An example of interhelical knob-in-hole packing. The knob is in yellow and the hole in blue.

more interactions of these types or additional geometric conditions (as is the case, e.g., for a hydrogen bond), the initial conformations can be checked and discarded if they do not meet these additional conditions. For our applications, four interactions would tend to constrain the conformations well.

For analysis of packing, we consider knob-in-hole interactions. Each knob is a reduced rigid representative of a side chain based on amino acid type, and each hole is a predefined region (Fig. 2). Each hole is given by the space between a set of helix

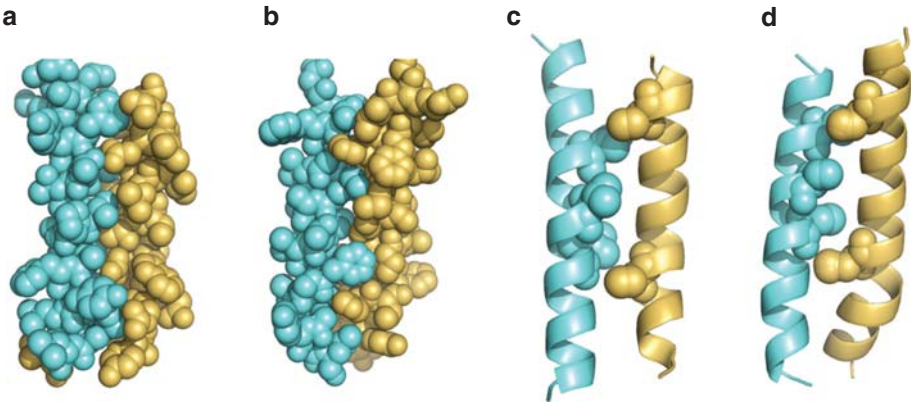


**Fig. 3.** **Equivalently packed helices.** These unrelated helix pairs have very similar folds. Fully spacefilled models: (**a**) helix pair from bR, (1C3W); (**b**) helix pair from cytochrome B6-F (1Q90). Note how different the side chains are in panels (**a**, **b**). In panels (**c**, **d**) we see the common packing of knobs-in-holes in the two structures despite their sequence differences. The knobs are fixed reduced representatives of side chains based on residue type. The five knobs shown spacefilled are packed in a nearly equivalent way in (**c**) the helix pair from 1C3W, also shown in (**a**, **d**) the helix pair from 1Q90, also shown in (**b**).

residues $i$, $i+3$, $i+4$, $i+7$, but subdivided into three regions (Harrington and Ben-Tal 2009). Four knobs-in-holes would usually constrain a conformation of a helix pair quite well. But the knobs of different amino acids are similar enough so that this is often true even when comparing the packing of helices of different amino acid composition. That is, if four corresponding knobs pack into the same four corresponding holes, the two helix pairs will usually have similar conformations, especially around this region. In Fig. 3, we see unrelated helix pairs with five knobs packed in the same way in both pairs. This can be seen in the helix pair clusters, except there is no condition that knobs pack into corresponding positions on the two pairs of helices, and so the result is local to the packing region.

### 3.9 Iterative reassembly of full TM helix bundles using interactions of the five types

To analyze the interactions of the five types as constraints, we used rigid motions to iteratively reassemble the helix bundle backbone of each protein using only its set of the five types of interhelical interactions, predefined interaction geometries, and individual helix backbones. Beginning with N rigid separate pieces, initially the individual helices, we fit two together and so obtained a new set of N–1 rigid pieces. After repeating this N–1 times, there is one piece at the end, the assembled structure (actually an ensemble of structures as explained above).

From each solved structure, the backbone conformations of the individual helices and the side chain conformations of those residues with a side chain atom explicitly in an interhelical interaction of one of the five types were taken. Those native side chain conformations are fixed and rigid. The side chains were not taken for residues in the two packing interactions. All other side chains had a fixed, rigid, reduced representation based on residue type that is intended to give the obstruction created by a side chain of that residue type irrespective of rotameric state; they were not derived from the native structures.

The scoring of the structures depended only on overlap penalties and the geometric conditions imposed by the interactions. It does not approximate energy: in particular, a VDW term was not used. For details, see Harrington and Ben-Tal (2009).

The order of reassembly was chosen to mimic a plausible translocon-guided folding pathway for each protein, and so we attempted to rebuild the structures from the N-terminus in a sequence order preserving fashion.

For example, here is our iterative assembly for the voltage sensor, 1ORS. At each step, we put two rigid pieces together using the interhelical interactions between them to produce a new fixed piece (actually an ensemble as described before), as shown in Fig. 4. We first assemble the first two helices, then add 3-a to the single

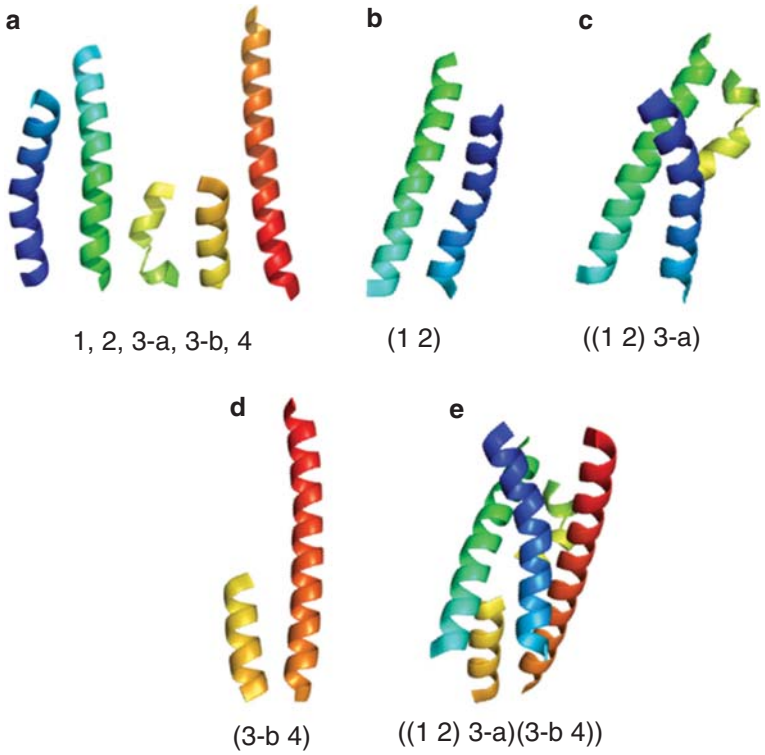


**Fig. 4. Assembly order for voltage sensor.** An example of our iterative sequence order respecting assembly for the voltage sensor 1ORS. We begin with the individual helices 1, 2, 3-a, 3-b, 4, and assemble iteratively in the order shown. (**a**) The helices to be assembled in sequence order. (**b**) The first two helices assembled, (1 2). (**c**) The third (half) helix is assembled with the first two, ((1 2) 3-a). (**d**) The last two helices assembled together, (3-b 4). (**e**) The piece made up of the first three helices and the piece made up of the last two are assembled together to build the full structure, (((1 2) 3-a) (3-b 4)). Figure adapted from Harrington and Ben-Tal (2009).

piece (1 2). At this point, there are insufficiently many interactions to add 3-b to the first piece ((1 2) 3-a), so we next assemble 3-b and 4, and finally put ((1 2) 3-a) and (3-b 4) together to obtain the full structure (((1 2) 3-a) (3-b 4)).

### 3.10 The sets of the five types of particularly favorable interactions determine the packing of helices in the native structures of a diverse test set

For a diverse test set of 15 TM helical proteins, the structures rebuilt in the fashion outlined above had an average ensemble-average Cα rmsd from the native of 1.03 Å

(Harrington and Ben-Tal 2009). Furthermore, with the exception of aquaporin, the structures could be rebuilt in a sequence order preserving fashion consistent with translocon-aided folding. In the case of aquaporin, the half-helices needed to be assembled slightly out of order. This might relate to experimental results indicating some unusual insertion behavior of half-helices in general and helices in other aquaporins in particular (Pitonzo and Skach 2006; Jaud et al. 2009).

Determining sets of interactions of the five types seem to be a very common structural feature of solved TM helical proteins, and for good physical reasons as we will explain. But there are proteins without them (e.g., proteins with large prosthetic groups). For a discussion, see Harrington and Ben-Tal (2009).

### 3.11 Distribution of particularly stabilizing residues, folding funnels, and the construction of low-energy minima

If the residues participating in these types of interactions are likely to make particularly favorable contributions to stability, then their distribution in determining sets of interactions partially explains how sequence specifies structure. These sets of interactions help to create low-energy minima for two reasons. First, the abundance of these particularly favorable interactions would tend to act to make the structure a low-energy one. Second, when the backbone positions of the helices are significantly perturbed, some of the determining set of the interactions will necessarily be broken. At the very least and for very few perturbations, some side chains must be flipped and so rotameric barriers crossed. If we assume the interaction energies are strong enough, it will be difficult to compensate for the lost interactions of the five types given their geometric and partner specificity and the rarity of possible participants. Thus the energies of the perturbed structures would tend to be higher. In contrast, one could not usually say the same of a "determining set of VDW interactions" because of the density and promiscuity of VDW interactions.

The iterative assembly (consistent with translocon-aided folding) and the determining sets of interactions can also be seen as a geometric recipe for creating folding funnels. The interactions of the five types are supposed to be individually and locally superior to generic contacts and so can successively funnel and collectively trap the native backbone. That this could be done in a controlled iterative way aided by the translocon makes the process much simpler. The native backbone conformations (and sub-conformations) usually have many interactions of the five types that can stabilize them in addition to the ones that appear in the solved structures since side chains that participate in these interactions can adopt different conformations and form different interactions. These additional interactions could further aid the funneling process.

### 3.12 Cooperativity with packing

Conformations of proteins are restricted by simple packing rules of the type we have seen for helix pairs. Between every pair of pieces (helices or subbundles) we used for reassembly, there must be at least four interactions of the five types. But if we instead consider either knob-in-hole or aromatic interactions (for loosely packed proteins) between those same pieces, we find that there are also at least three of these types of packing interactions (overwhelmingly knob-in-hole; Harrington 2009, unpublished data). Thus these native conformations are also quite constrained by simple packing interactions (although not as constrained as by the interactions of the five types).

### 3.13 Static structures versus ensembles

The sets of interactions of the five types were derived from crystal structures. In reality, an ensemble of structures underlies any crystal structure, which complicates the analysis. Due both to the limits of resolution and the underlying multiplicity of structures, it can be difficult to read the set of these interactions from a structure. To deal with this, we added error terms to the geometric conditions; to fully address it would require native state ensembles for all of the crystal structures. But the fact that for a native backbone conformation there can be multiple determining sets due to, e.g., bond switching (residues in hydrogen bonds changing partners or rotamers) does not contradict our analysis. If a determining set of interactions is fixed, the backbone conformations are highly constrained but can slightly jitter without breaking the interactions. If one began with a different determining set for the same backbone conformation, the two backbone ensembles constrained by the two different determining sets would not be identical, but very similar.

## 4 Conservation and diversity of determining sets of stabilizing interactions

If such structural importance is attributed to the determining sets of interactions of these types, then what happens in proteins of the same fold? For homologs, do the residues participating in these kinds of interactions have to be conserved or is some diversity possible? What about for remote homologs or proteins with related folds? How many ways are there to fix similar backbone conformations using these types of interactions?

As we have seen for helix pairs, great diversity in the determining sets of interactions is possible for very similar backbone conformations from unrelated proteins (Fig. 1). In homologs, we can also see this diversity.

For homologs, the conservation pattern of residues participating in the determining sets of interactions is mixed. In general, buried residues are more conserved than exposed ones (Baldwin et al. 1997; Fleishman et al. 2004; Liao et al. 2005;

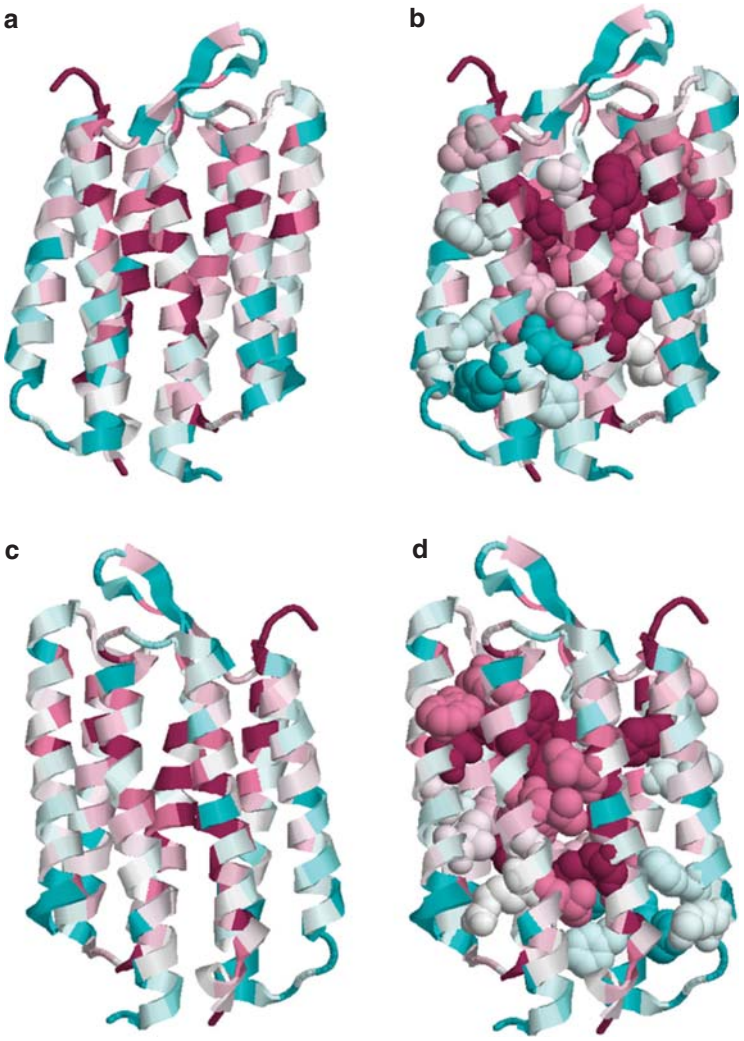


**Fig. 5. Conservation of residues in bR and its determining set.** The conservation is indicated by color from turquoise to maroon: turquoise means highly variable, and maroon means highly conserved. White is intermediate. (**a**) One view of bR. (**b**) Here the residues involved in the five types of interactions are spacefilled. (**c**) Another view of bR. (**d**) Another view of the residues in the determining set as spacefilled. Thus almost all but the most exposed residues in the determining set of interactions are conserved. The exposed ones can vary.

Park and Helms 2007; Hildebrand et al. 2008). For helical membrane proteins, it has been found that residues in hydrogen bonds are 40% more conserved than exposed residues while buried residues are 25% more conserved than exposed residues (Hildebrand et al. 2008). For many closely packed small residues, substitution by a bulkier residue would require a substantial change in the backbone conformation, and so such residues will tend to be conserved in the same fold for that reason alone. Buried residues in the determining sets tend to be very highly conserved. For more exposed interactions in the determining sets, diverse alternative sets of these constraints are commonly seen, and such residues less conserved (Liu et al. 2004). We will now look at this phenomenon in bR and its homologs.

### 4.1 Conservation and diversity of the determining sets of interactions of bR

We will consider the two most conserved categories of residues as classified by ConSurf: highly conserved and conserved (Fig. 5; Glaser et al. 2003). In the helical part of bR, there are 17 residues within the inferred hydrocarbon region classified as highly conserved (the top category), and 11 of these participate in the determining set of interactions. Of the six remaining, two are helix-kinking prolines, one is K216, which is critical for binding retinal, and one is M60, which we have discussed as a guide for Y57. There are 18 helix residues classified as conserved (the next-to-top category), 11 of which are part of the determining set of interactions. Again, many of those conserved residues not participating in the determining set of interactions are known to be functionally important or related to important secondary structure features.

The less buried residues in the determining sets are much less conserved. In the crystal structures of the homologs of bR, we can see the diversity of the less buried parts of the determining sets of interactions (Fig. 6). Intuitively, it is unsurprising that mutations of more exposed residues are more likely to result in diverse determining sets of interactions simply because there would not be the additional geometric constraints imposed by surrounding residues, so the evolutionary process would not require as many concerted mutations.

## 5 Determining sets, multiple states, and motion

Determining sets of interactions highly constrain the positions of helix bundle backbones, so how do they relate to the dynamic properties of TM helical proteins? The simplest answer according to the constraint philosophy is that each conformational state (or highly constrained ensemble) corresponds to a different determining set,

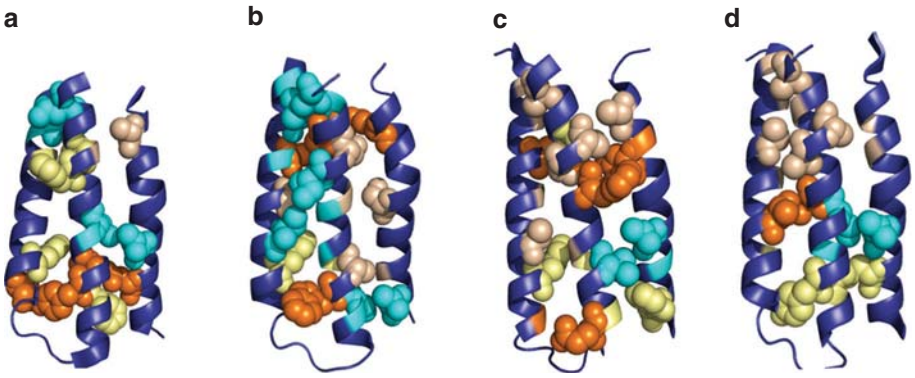


**Fig. 6. Comparison of homologs: The same fold with diverse determining sets of interactions.** The residues in the determining sets are spacefilled and color-coded as in Fig. 1. The panels (**a**, **b**) show bacteriorhodopsin and halorhodopsin. (**a**) Bacteriorhodopsin (1C3W). (**b**) Halorhodopsin (1E12). The panels (**c**, **d**) show sensory rhodopsins II. (**c**) Sensory rhodopsin II, Anabaena (1XIO). (**d**) Sensory rhodopsin II, N. pharaonis (1H68). Figure adapted from Harrington and Ben-Tal (2009).

and motion would result from a collection of these states. There should also be a reasonable transition between states, which from the constraint perspective would mean a way to switch smoothly between the different determining sets of interactions corresponding to the two states. Suggested mechanisms for these transitions include deformation of hydrogen bonds, with non-native bonds along the transitions or rotameric flips, and sliding in interfaces dominated by small residues (Perozo et al. 2002; Curran and Engelman 2003; Hildebrand et al. 2008; Gardino et al. 2009).

### 5.1 Multiple states and motion in the ErbB family

The epidermal growth factor family of receptor tyrosine kinases is an interesting case in point. The members of this family (ErbB1, ErbB2, ErbB3, and ErbB4) play critical roles in a variety of physiological processes and their malfunction has been associated with many cancers. Each member has the same overall components: an extracellular ligand-binding domain connected to a single TM helix, which is connected to a cytoplasmic tyrosine kinase domain. Usually the formation of hetero- and homodimers of the TM helices in this family is induced by the binding of ligands to the extracellular domain. A ligand specific to ErbB2 has not been found, but it can be affected by the ligand-binding of the other family members. Furthermore, it does appear to form active homodimers on its own. Dimerization can trigger the tyrosine kinase activity of the cytoplasmic domain, and so it has a crucial functional role for these receptors.


**Fig. 7. Multiple sequence alignment around the TM segments of the four human ErbB paralogs.** The N-terminal GxxxG-like motif is indicated in yellow and the C-terminal GxxxG-like motif in blue. Note that ErbB3 does not have a C-terminal motif of this type.

The TM dimers of this family make attractive candidates for study because of their simplicity and medical significance. The TM helices all contain similar dimerization motifs near the N-terminus of their helices similar to GxxxG, but actually defined earlier (Sternberg and Gullick 1990). Seven residues later in their helix sequences, for all but ErbB3, there appears to be another dimerization motif related to the famous GxxxG (Fig. 7). This suggests that there are at least two main conformational states in the homo and heterodimers: one corresponding to the N-terminal motif and the other to the C-terminal motif. Using a scoring function based on rewarding the packing of small residues and a grid search for the helix pair conformations, Fleishman et al. (2002) produced model structures for these two states for ErbB2 and proposed a molecular switch for activation based on them. In this switch model, the conformation induced by the N-terminal motif is the active form of ErbB2, and the conformation induced by the C-terminal motif the inactive form. This model was found to explain some known disease-causing mutations in detail, and later the pathway between them was validated by motion-planning methods (Fleishman et al. 2002; Enosh et al. 2007). Both these proposed conformations, as well as the proposed pathway between them, are consistent with the determining set perspective. The two GxxxG-like motifs as well as the nearby polar (N-terminal), aromatic (C-terminal), and V/I/L/T residues can all participate in the five types of interactions and highly constrain the conformations. The proposed intermediate states feature extensive close knob-in-hole V/I/L/T patch packing and some small residue packing, and so are also in line with the determining set philosophy of motion.

These ideas have experimental support. Solution NMR structures for the TM dimer of ErbB2 have been found, and they closely agree with the earlier proposed model for the active state (Bocharov et al. 2007). The conformations use the N-terminal dimerization motif as expected, and there are interhelical hydrogen bonds between the hydroxyl groups of Ser656 and between the hydroxyl groups of Thr652. These interhelical hydrogen bonds are transient and vary among the conformations in the ensemble (Fig. 8). Additionally, an aromatic–aromatic edge-to-face interhelical in-

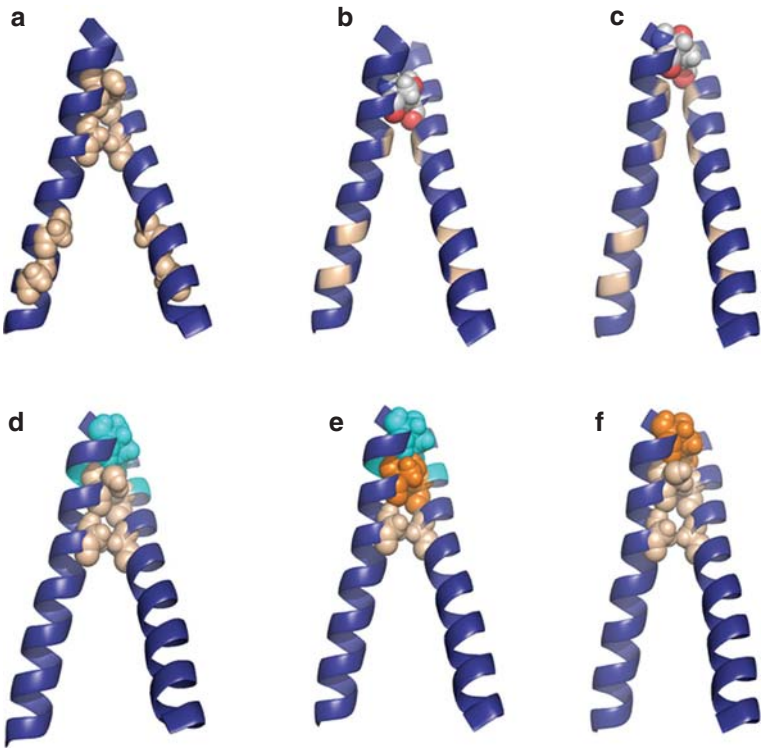


**Fig. 8. NMR structures of the active state of ErbB2.** The stabilizing interactions for the ensemble of the active state of the ErbB2 dimer (2JWA). For both upper and lower panels, three different models from the ensemble are shown. Upper panel: (**a**) shows the two GxxxG-like dimerization motifs spacefilled and colored in wheat. The unused C-terminal motif is believed to mediate dimerization of the inactive state. (**b**) Transient hydrogen bonds between the hydroxyl groups of S656 are seen; here is one displayed spacefilled. (**c**) Transient hydrogen bonds between the hydroxyl groups of T652 are seen; here one is shown spacefilled. Lower panel: Since the hydrogen bonds vary among the ensemble members, their determining sets differ. However, both S656 and T652 are in particularly stabilizing knob-in-hole packing interactions whether or not they hydrogen bond, so the conformations are well-constrained even without the hydrogen bonds. The determining sets of interactions are color-coded as described for Fig. 1. (**d**) Neither S656 nor T652 form hydrogen bonds in this conformation, so there is only small residue packing and I//V/L/T packing. (**e**) There is a S656–S656 hydrogen bond shown in orange in this conformation. (**f**) There is a T652–T652 bond shown in orange in this conformation.

teraction is reported. All of this is consistent with the determining set perspective we have described.

There is also experimental support for the proposed inactive state of ErbB2. Many experiments have tested the suspected dimerization motifs in the ErbB family. Escher et al. (2009) studied the dimerization capabilities in a biological membrane of both N- and C-terminal dimerization motifs for both homo- and heterodimers.

The N- and C-terminal motifs were studied separately by dividing each TM helix into two parts to create two new TM segments. One of the new TM segments contained the N-terminal motif only and the other contained the C-terminal motif only, such that the two motifs occupied equivalent positions on their respective segments. For ErbB2, it was found that these C-terminal motif segments dimerized, supporting this putative dimerization motif and hence also the putative inactive state structure.

There is very active study of the entire ErbB family; for reviews, see (Landau and Ben-Tal 2008; Hynes and MacDonald 2009; Lemmon 2009). One can hope that analysis of the type done successfully for ErbB2 will also work for the other possible states of the homo- and heterodimers of this family. Perhaps such methods can also be applied to other tyrosine kinase receptors.

## 6 Conclusion

From the earliest days of structural biology, the notion of energetically important interactions as geometric constraints has been key to many classic discoveries. We have argued that the idea also sheds light on the complex problems of understanding TM helical proteins today. The modern emphasis on multiple states and ensembles of structures as the key to function (Henzler-Wildman and Kern 2007) has a direct connection to old-fashioned model building. Just as model-building was guided by physically important interactions as constraints, these multiple states and the transitions between them can be seen in part as geometrically created and mediated by determining sets of well-known favorable interactions. The folding and dynamics of TM helical proteins, their structure prediction and eventual design can all be seen more clearly in this light, despite our imperfect knowledge of these stabilizing interactions, especially their energetics. We believe that these ideas form the foundation for new top–down algorithms for structure and motion prediction which could bridge the gap to more detailed bottom–up approaches such as molecular dynamics simulations.

## References

Albert AD and Litman BJ (1978) Independent structural domains in the membrane protein bovine rhodopsin. Biochemistry 17: 3893–3900

Arbely E and Arkin I (2004) Experimental measurement of the strength of a Cα–H–O bond in a lipid bilayer. J Am Chem Soc 126: 5362-5363

Baldwin JM, Schertler GF, Unger VM (1997) An α-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. J Mol Biol 272: 144–164

Bocharov EV, Mineev KS, Volynsky PE, et al. (2007) Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. J Biol Chem 283: 6950–6956

Bogan AA and Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280: 1–9
Bogdanov M, Xie J, Heacock P, et al. (2008) To flip or not to flip: lipid–protein charge interactions are a determinant of final membrane topology. J Cell Biol 182: 925–935
Bond PJ, Holyoake J, Ivetac A, et al. (2007) Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J Struct Biol 157: 593–605
Bowie JU (2005) Solving the membrane protein folding problem. Nature 438: 581–589
Chamberlain AK, Lee Y, Kim S, et al. (2004) Snorkeling preferences foster an amino acid composition bias in transmembrane helices. J Mol Biol 339: 471–479
Chiang CS, Shirinian L, Sukharev S (2005) Capping transmembrane helices of MscL with aromatic residues changes channel response to membrane stretch. Biochemistry 44: 12589–12597
Curran AR and Engelman DM (2003) Sequence motifs, polar interactions and conformational changes in helical membrane proteins. Curr Opin Struct Biol 13: 412–417
Crick F (1953) The packing of α-helices: simple coiled-coils. Acta Crysta 6: 689–697
de Planque MRR, Bonev BB, Demmers JAA, et al. (2003) Interfacial anchor properties of tryptophan residues in transmembrane peptides can dominate over hydrophobic matching effects in peptide-lipid interactions. Biochemistry 42: 5341–5348
Dougherty D (1996) Cation-pi interactions in chemistry and biology: a new view Benzene, Phe, Tyr, and Trp. Science 271: 163–168
Enosh A, Fleishman SJ, Ben-Tal N, et al. (2004) Assigning transmembrane segments to helices in intermediate-resolution structures. Bioinformatics 20: i122–i129
Enosh A, Fleishman SJ, Ben-Tal N, et al. (2007) Prediction and simulation of motion in pairs of transmembrane alpha-helices. Bioinformatics 23(2): e212–e218
Escher C, Cymer F, Schneider D (2009) Two GxxxG-like motifs facilitate promiscuous interactions of the human ErbB transmembrane domains. J Mol Biol 389: 10–16
Faham S, Yang D, Bare E, et al. (2004) Side-chain contributions to membrane protein structure and stability. J Mol Biol 335: 297–305
Fleming KG and Engelman DM (2001) Specificity in transmembrane helix–helix interactions can define a hierarchy of stability of sequence variants. Proc Natl Acad Sci USA 98: 14340–14344.
Fleishman SJ, Schlessinger J, Ben-Tal N (2002) A putative molecular-activation switch in the transmembrane domain of erbB2. Proc Natl Acad Sci USA 99: 15937–15940
Fleishman SJ, Harrington S, Friesner RA, et al. (2004) An automatic method for predicting transmembrane protein structures using cryo-EM and evolutionary data. Biophys J 87: 3448–3459
Gao J and Li Z (2009) Comparing four different approaches for the determination of inter-residue interactions provides insight for the structure prediction of helical membrane proteins. Biopolymers 91: 547–556
Gardino A, Villali J, Kivenson A, et al. (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. Cell 139: 1109–1118
Glaser F, Pupko T, Paz I, et al. (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. Bioinformatics 19: 163–164
Gratkowski H, Lear JD, DeGrado WF (2001) Polar side chains drive the association of model transmembrane peptides. Proc Natl Acad Sci USA 98: 880–885
Gurezka R, Laage R, Brosig B, Langosch DA (1999) Heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. J Biol Chem 274: 9265–9270
Harrington SE and Ben-Tal N (2009) Structural determinants of transmembrane helical proteins. Structure 17: 1092–1103

Hedin L, Ojemalm K, Bernsel A, et al. (2010) Membrane insertion of marginally hydrophobic transmembrane helices depends on sequence context. J Mol Biol 396: 221–229
Henzler-Wildman K and Kern D (2007) Dynamic personalities of proteins. Nature 450: 964–972
Hessa T, Kim H, Bihlmaier K, et al. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. Nature 433: 377–381
Hessa T, Meindl-Beinker NM, Bernsel A, et al. (2007) Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 450: 1026–1030
Hildebrand PW, Gunther S, Goede A, et al. (2008) Hydrogen-bonding and packing features of membrane proteins: functional implications. Biophys J 94: 1945–1953
Holt A and Killian JA (2009) Orientation and dynamics of transmembrane peptides: the power of simple models. Eur Biophys J 39: 609–621
Honig BH and Hubbell WL (1984) Stability of "salt bridges" in membrane proteins. Proc Natl Acad Sci USA 81: 5412–5416
Huang KS, Bayley H, Liao MJ, et al. (1981) Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteo-lytic fragments. J Biol Chem 256: 3802–3809
Hynes NE and MacDonald G (2009) ErbB receptors and signaling pathways in cancer. Curr Opin Cell Biol 21: 177–184
Jaud S, Fernandez-Vidal M, Nilsson I, et al. (2009) Insertion of short transmembrane helices by the Sec61 translocon. Proc Natl Acad Sci USA 106: 11588–11593
Joh NH, Min A, Faham S, et al. (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. Nature 453: 1266–1270
Johnson RM, Hecht K, Deber CM (2007) Aromatic and cation–pi interactions enhance helix–helix association in a membrane environment. Biochemistry 46: 9208–9214
Kida Y, Morimoto F, Sakaguchi M (2007) Two translocating hydrophilic segments of a nascent chain span the ER membrane during multispanning protein topogenesis. J Cell Biol 179: 1441–1452
Killian JA and von Heijne G (2000) How proteins adapt to a membrane-water interface. Trends Biochem Sci 25: 429–434
Krishnakumar SS and London E (2007) Effect of sequence hydrophobicity and bilayer width upon the minimum length required for he formation of transmembrane helices in membranes. J Mol Biol 374: 671–687
Landau M and Ben-Tal N (2008) Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in ErbB receptors. Biochim Biophys Acta 1785: 12–31
Landin JS, Katragadda M, Albert AD (2001) Thermal destabilization of rhodopsin and opsin by proteolytic cleavage in bovine rod outer segment disk membranes. Biochemistry 40: 11176–11183
Lemmon MA (2009) Ligand-induced ErbB receptor dimerization. Exp Cell Res 15: 638–648
Liao H, Yeh W, Chiang D, et al. (2005) Protein sequence entropy is closely related to packing density and hydrophobicity. Protein Eng Des Sel 18: 59–64
Liao MJ, Huang KS, Khorana HG (1984) Regeneration of native bacteriorhodopsin structure from fragments. J Biol Chem 259: 4200–4204
Litman BJ (1979) Rhodopsin: its molecular substructure and phospholipid interactions. Photochem Photobiol 29: 671–677
Liu W, Eilers M, Patel AB, et al. (2004) Helix packing moments reveal diversity and conservation in membrane proteins. J Mol Biol 337: 713–729

Lundbaek JA, Collingwood SA, Ingolfsson HI, et al. (2010) Lipid bilayer regulation of membrane protein function: gramicidin channels as molecular force probes. J R Soc Interface 7: 373–395

Marti T (1998) Refolding of bacteriorhodopsin from expressed polypeptide fragments. J Biol Chem 273: 9312–9322

Nanda V and Schmiedekamp A (2008) Are aromatic carbon donor hydrogen bonds linear in proteins? Proteins 70: 489–497

Pabuwal V and Li Z (2008) Network pattern of residue packing in helical membrane proteins and its application in membrane protein structure prediction. Prot Eng Des Sel 21: 55–64

Pabuwal V and Li Z (2009) Comparative analysis of the packing topology of structurally important residues in helical membrane and soluble proteins. Prot Eng Des Sel 22: 67–73

Park Y and Helms V (2007) On the derivation of propensity scales for predicting exposed transmembrane residues of helical membrane proteins. Bioinformatics 23: 701–708

Perozo E, Kloda A, Cortes DM, et al. (2002) Physical principles underlying the transduction of bilayer deformation forces during mechanosensitive channel gating. Nat Struct Biol 9: 696–703

Pitonzo D and Skach WR (2006) Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. Biochim Biophys Acta 1758: 976–988

Pitonzo D, Yang Z, Matsumura Y, et al. (2009) Sequence-specific retention and regulated integration of a nascent membrane protein by the endoplasmic reticulum Sec61 translocon. Mol Biol Cell 20: 685–698

Popot JL and Engelman DM (1990) Membrane protein folding and oligomerization: the two-stage model. Biochemistry 29: 4031–4037

Popot JL, Gerchman SE, Engelman DM (1987) Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. J Mol Biol 198: 655–676

Ren J, Lew S, Wang J, et al. (1999) Control of the transmembrane orientation and interhelical interactions within membranes by hydrophobic helix length. Biochemistry 38: 5905–5912

Ridge KD, Lee SS, Yao LL (1995) In vivo assembly of rhodopsin from expressed polypeptide fragments. Proc Natl Acad Sci USA 92: 3204–3208

Russ WP and Engelman DM (2000) The GxxxG motif: a framework for transmembrane helix–helix association. J Mol Biol 296: 911–919

Sadlish H, Pitonzo D, Johnson AE, et al. (2005) Sequential triage of transmembrane segments by Sec61α during biogenesis of a native multispanning membrane protein. Nat Struct Mol Biol 12: 870–878

Sal-Man N, Gerber D, Bloch I, et al. (2007) Specificity in transmembrane helix–helix interaction mediated by aromatic residues. J Biol Chem 282: 19753–19761

Schneider D and Engelman DM (2004) Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. J Mol Biol 343: 799–804

Senes A, Engel DE, DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr Opin Struct Biol 14: 465–479

Senes A, Ubarretxena-Belandia I, Engelman DM (2001) The Calpha–H⋯O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. Proc Natl Acad Sci USA 98: 9056–9061

Skach WR (2007) The expanding role of the ER translocon in membrane protein folding. J Cell Biol 179: 1333–1335

Sternberg MJ and Gullick WJ (1990) A sequence motif in the transmembrane region of growth factor receptors with tyrosine kinase activity mediates dimerization. Prot Eng 3: 245–248

Strandberg E, Morein S, Rijkers DTS, et al. (2002) Lipid dependence of membrane anchoring properties and snorkeling behavior of aromatic and charged residues in transmembrane peptides. Biochemistry 41: 7190–7198

Tastan O, Klein-Seetharaman J, Meirovitch H (2009) The effect of loops on the structural organization of a-helical membrane proteins. Biophys J 96: 2299–2312

van Klompenburg W, Nilsson I, von Heijne G, et al. (1997) Anionic phospholipids are determinants of membrane protein topology. EMBO J 16: 4261–4266

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

Walters RFS and DeGrado WF (2006) Helix-packing motifs in membrane proteins. Proc Natl Acad Sci USA 103: 13658–13663

White SH and von Heijne G (2008) How translocons select transmembrane helices. Annu Rev Biophys 37: 23–42

Zhang Y, Kulp DW, Lear JD, et al. (2009) Experimental and computational evaluation of forces directing the association of transmembrane helices. J Am Chem Soc 131: 11341–11343

Zhou FX, Cocco MJ, Russ WP, et al. (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat Struct Biol 7: 154–160