

Analysis and Prediction of Helix–Helix Interactions in Membrane Channels and Transporters

Peter Werner Hildebrand,^{1*} Stephan Lorenzen,² Andrean Goede,¹ and Robert Preissner¹

¹Institute of Biochemistry Charité, Universitätsmedizin Berlin, Monbijoustraße 2, 10117 Berlin, Germany

²Bioinformatics Center and Department of Molecular Bioscience, University of Kansas, Lawrence, Kansas

ABSTRACT Membrane proteins span a large variety of different functions such as cell-surface receptors, redox proteins, ion channels, and transporters. Proteins with functional pores show different characteristics of helix–helix packing as other helical membrane proteins. We found that the helix–helix contacts of 13 nonhomologous high-resolution structures of membrane channels and transporters are mainly accomplished by weakly polar amino acids (G > S > T > F) that preferably create contacts every fourth residue, typical for right-handed helix crossings. There is a strong correlation between the now available biological hydrophobicity scale and the propensities of the weakly polar and hydrophobic residues to be buried at helix–helix interfaces or to be exposed to the lipids in membrane channels and transporters. The polar residues, however, make no major contribution towards the packing of their transmembrane helices, and are therefore subsumed to be primarily exposed to the polar milieu during the folding process. The contact formation of membrane channels and transporters is therefore ruled by the solubility of the residues, which we suppose to be the driving force for the assembly of their transmembrane helices. By contrast, in 14 nonhomologous high-resolution structures of other membrane protein coils, also large and polar amino acids (D > S > M > Q) create characteristic contacts every 3.5th residues, which is a signature for left-handed helix crossings. Accordingly, it seems that dependent on the function, different concepts of folding and stabilization are realized for helical membrane proteins. Using a sequence-based matrix prediction method these differences are exploited to improve the prediction of buried and exposed residues of transmembrane helices significantly. When the sequence motifs typical for membrane channels and transporters were applied for the prediction of helix–helix contacts the quality of prediction rises by 16% to an average value of 76%, compared to the same approach when only single amino acid positions are taken into account. Proteins 2006;64:253–262. © 2006 Wiley-Liss, Inc.

Key words: helix–helix interactions; membrane channels; transporters

INTRODUCTION

The exchange of information and substances between different cells and cellular compartments is entirely con-

trolled by integral membrane proteins. This protein class includes members with completely different functions, such as membrane channels and transporters (referred to as “channels”) or cell-surface receptors and redox proteins (referred to as “membrane coils”). Despite recent efforts,^{1,2} the difficulties in expression and crystallization still complicate the biophysical and structural characterization of membrane proteins.³ Thus, high-resolution structures are available for only about 50 different membrane proteins at the PDB.⁴ This is only a minor part of all membrane proteins, which take up to 30% of the genes of various genomes.⁵ Therefore, tertiary structure prediction methods are important in gaining information about the 3D structure of membrane proteins that are not yet crystallized.

The ongoing process in membrane protein research highlights the fact that the protein structure is predominantly stabilized by the membrane spanning regions, which are either of the all-alpha or the all-beta type.^{6–8} Beta-barrel membrane proteins (all-beta type) are exclusively embedded in prokaryotic lipid bilayers. Thus, the helix bundle membrane proteins (all-alpha type) are the predominant and bio-medically most interesting group. This structural variety, which is reduced in comparison to water-soluble globular proteins, is mainly caused by the nature of the lipid bilayer: Within its hydrophobic surrounding, all the hydrogen bonds need to be saturated. Thus, the transmembrane helices have to be completely folded before they are imbedded into the membrane.^{6,7}

In eukaryotic cells, most transmembrane helices are synthesized within the Ribosome-Translocon Complex before they are integrated into the membrane.⁹ The translocation machinery requires the amino acid composition of the polypeptide to be very hydrophobic, with polar amino acids being located at the helix termini. Now, in vitro measurements substantiate that polypeptide chains with periodically positioned polar amino acids at other positions of the helix are inserted into the lipid bilayer as well.¹⁰

Grant sponsor: Federal Ministry of Education and Research of Germany (BMBF); Grant sponsor: the Berlin Centre for Genome-Based Bioinformatics (BCB).

*Correspondence to: Peter Werner Hildebrand, Institute of Biochemistry Charité, Universitätsmedizin Berlin, Monbijoustraße 2, 10117 Berlin, Germany. E-mail: peter.hildebrand@charite.de

Received 2 November 2005; Revised 23 January 2006; Accepted 24 January 2006

Published online 22 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20959

However, the mechanism underlying the helix assembly is still under discussion.^{11–13} In channels, the rules determining the folding of the tertiary structure may again be very specific, because many residues are in contact with water filled pores or cavities.^{14,15} This fact may lead to a very typical amino acid composition of the contact sites of channels, which has not been quantified yet. For the enormous interest in the folding and prediction of their tertiary structure,^{16–20} the systematic analysis of helix–helix interactions in channels is an outstanding task.

Because of the restrictions imposed on the protein fold by the lipid bilayer, the topological space for the modeling of helical membrane proteins is generally smaller than for water-soluble globular proteins.²¹ However, this attempt is commonly hampered by the fact that the amino acid composition is nearly as hydrophobic in the interior as at the lipid exposed exterior of membrane proteins.^{22,23} In contrast to water-soluble globular proteins, hydrophobicity is therefore only a weak discriminator for the prediction of the exposed or buried faces of helices spanning the lipid bilayer. Thus, more specific criteria are needed to predict the different contact sites of transmembrane helices accurately.

The orientation of the helices within the lipid bilayer is determined by various factors as the amino acid composition of the helix caps and by short loops connecting helix neighbors.^{24,25} Folding experiments, however, manifest that integral membrane proteins also reassemble from fragments, and that their folding and stability is largely determined by the amino acid composition of the transmembrane helices.²⁶ The size and chemistry of buried residues are conserved in sequence motifs, and hence, play a leading role in the assembly of transmembrane helices.^{18,27,28} The investigations into helix packing, however, evidently reveal that the interior of membrane proteins is not homogeneously packed.^{15,29–31}

We have previously shown that channels, which make up half of all known helical membrane protein structures, are packed more loosely than other helical membrane proteins.¹⁵ In particular, there is a frequent placement of polar side chains close to polar voids or polar pockets that are lined along the protein channel. These functionally important protein cavities have not been valued equivalently by previous methods, resulting in a general overestimation of the packing density of helical membrane proteins. The postulation that helical membrane proteins are packed more tightly as helices of globular proteins is consequently not verified by recent analysis.^{15,31} For that reason, it seems logically consistent that not all residues that are buried in the interior of helical membrane proteins essentially contribute to the assembly and stability of transmembrane helices.

It has been suggested that the geometric fit between two helical surfaces determines the specificity of helix–helix interactions, whereas hydrogen bonding is important for the stabilization of a preformed helix dimer.^{12,32} Small axial distances between transmembrane helices allow for the formation of backbone-to-backbone hydrogen bonds, which are supposed to be of considerable strength inside

the membrane.³³ This is highlighted by the fact that the near native conformation of such transmembrane helix pairs can be modeled when small residues are buried at helix–helix interfaces.^{34,35} This structural motif seems to be very characteristic for channels, indicating that these membrane proteins are stabilized by additional backbone-to-backbone hydrogen bonds.¹⁵

Most prediction methods use knowledge-based scales for the calculation of exposed or buried helix faces derived from sequential and structural data.^{36–38} Recently, the original idea³⁶ of combining conservation criteria with propensity scales to enhance the prediction of buried helix faces in membrane proteins has been reused.^{38,39} But for many membrane proteins no close relatives are known and multiple sequence alignments are not at hand. However, now there is sufficient structural data available to perform predictions on membrane proteins with different functions.

To improve the prediction of helix–helix contacts, we implemented a knowledge-based method that specifically uses the contact information derived from 13 channel structures and compare it to an equivalent scale derived from 14 membrane coils. To account for sequential patterns, the propensities of the residues next to a certain amino acid contact are also examined for the prediction. This is important because sequential motifs that are specific for right-handed helix–helix interactions are predominant in channels,^{18,25} which has not systematically been considered and applied for the 3D prediction yet. The results offer new insights into the mechanisms underlying the folding of helical membrane proteins.

METHODS

Data Set

For the regulated opening or closing of the pore underlying their function, membrane channels and transporters are subsumed here under the term “channels.” In contrast, proton pumps, receptors, and photosystems are classified as “membrane coils,” because the helix interaction motifs are similar to those characteristic for soluble coiled coils, with predominantly left-handed helix–helix crossings.²⁵ To be comparable to another analysis³⁹ the nonredundant data set (proteins with resolution >3.5 Å and sequence identity >30% within the transmembrane domains were excluded) of our previous analysis was updated (Table I).

Definition of Contact Types

The residues of neighboring helices contact each other when the atomic van der Waals surfaces are closer than a given cutoff distance. The size and atomic composition of a molecular surface patch varies with the chosen cutoff distance.⁴⁰ Hence, it is important that the same measure is used when contacts of different data sets are compared.⁴¹ Accordingly, a distance cutoff value of 1.5 Å between the atomic van der Waals surfaces of two neighboring helices was used for the calculation of helix–helix patches, which represents the nature of contact regions of helical membrane proteins.¹⁵

For the definition of contact types hetero atoms were not taken into account. Lipids and bulk solvent are only rarely

TABLE I. Prediction of Buried (Channels) and Exposed (Membrane Coil) Residues, Using Two Different Propensity Scales

PDB code	Protein name	N helices	Resol. (Å)	AUC (channels-scale)	AUC (membrane-coils scale)
Channels (prediction of buried amino acids)					
<i>2bl2</i>	Rotor V-type ATPase	4 (4)	2.1	0.83	0.53
<i>leul</i>	Calcium ATPase	10 (9)	2.6	0.62	0.61
<i>lfx8</i>	Glycerol facilitator	8 (6)	2.2	0.70	0.45
<i>liwg</i>	AcrB multidrug efflux transporter	12 (11)	3.5	0.76	0.64
<i>lj4n</i>	AQP1 aquaporin water channel	8 (6)	2.2	0.72	0.67
<i>ljvm</i>	KcsA potassium channel	2 (2)	2.8	0.81	0.59
<i>lkpl</i>	CIC chloride channel	14 (14)	3.0	0.73	0.71
<i>ll7v</i>	BtuCD vitamin B ₁₂ transporter	10 (6)	3.2	0.77	0.65
<i>lmsl</i>	MscL mechanosensitive channel	2 (2)	3.5	0.71	0.55
<i>lokc</i>	Mitochondrial ADP/ATP carrier	6 (6)	2.2	0.63	0.63
<i>lpv7</i>	Lactose permease	12 (12)	3.5	0.75	0.53
<i>lpw4</i>	GlpT glycerol-3-phosphate transporter	12 (12)	3.3	0.66	0.61
<i>lrh5</i>	SecYEβ protein-conducting channel	12 (11)	3.2	0.69	0.67
<i>all</i>		102 (91)	2.9	0.72	0.60
Membrane-coils (prediction of exposed amino acids)					
<i>lar1</i>	Cytochrome <i>c</i> oxidase (<i>Paracoccus denitrificans</i>)	12 (11)	2.7	0.53	0.65
<i>lc3w</i>	Bacteriorhodopsin	7 (7)	1.6	0.43	0.67
<i>lel2</i>	Halorhodopsin	7 (7)	1.8	0.56	0.67
<i>leys</i>	Cytochrome <i>bc</i> ₁ complex	11 (11)	2.3	0.55	0.65
<i>lf88</i>	Rhodopsin	7 (5)	2.8	0.53	0.65
<i>lfum</i>	Fumarate reductase (<i>E. coli</i>)	5 (5)	3.3	0.67	0.70
<i>lh2s</i>	Sensory rhodopsin	9 (9)	1.9	0.59	0.71
<i>ljb0</i>	Photosystem I	33 (30)	2.5	0.60	0.71
<i>lkqf</i>	Formate dehydrogenase-N	5 (5)	1.6	0.41	0.73
<i>lnek</i>	Succinate dehydrogenase	6 (6)	2.6	0.52	0.83
<i>2occ</i>	Cytochrome <i>c</i> oxidase (<i>Bos taurus</i>)	21 (21)	2.3	0.54	0.69
<i>lppj</i>	Photosynthetic reaction center	11 (10)	2.2	0.59	0.73
<i>lq16</i>	NarGHI nitrate reductase A	5 (5)	1.9	0.54	0.51
<i>lqla</i>	Fumarate reductase (<i>Wolinella succinogenes</i>)	5 (5)	2.2	0.62	0.76
<i>all</i>		130 (124)	2.2	0.54	0.69

The columns include the PDB code, the name of the protein, the number of helices used for the analysis, and prediction (helices shorter than four windings are not predicted), the crystallographic resolution of the structure, the AUC value using the channels scale and the AUC value using the membrane-coils scale.

reported in the PDB entries. Thus, the surface contacts are defined as all those helix atoms touched by a 1.4 Å radius probe rolling along the van der Waals spheres at the whole protein surface (Connolly surface). To differentiate between those atoms located within the lipophilic surrounding (helix–membrane) and those in contact with the polar milieu (helix–water), two parallel planes were drawn to isolate the expansions of the hydrophobic part of the lipid bilayer, using the criteria described in ref. 25.

At a defined cutoff distance the atoms of an amino acid may contribute to different contact types (e.g., when a residue is only partially exposed to the solvent). Thus, a second measure was applied that is defined by the minimum fraction of atoms that are either buried (helix–helix contacts) or solvent exposed (helix–membrane and helix–water contacts). Residues with a minimum fraction of 30% of their atoms being in contact with the solvent or with another helix are assigned as exposed or buried, respectively.

In this manner, about 70% of the residues located in the hydrophobic part of the membrane are unequivocally classified either as buried (~45%) or as lipid-exposed

(~25%). Moreover, another 20% of the residues of the transmembrane spanning region are partially buried and cannot be unequivocally assigned. Finally, about 10% of the residues are not assigned to any contact type. These residues are located in internal protein cavities or are lined along the channel. Because there are only few data on residues that contact the polar milieu via the channels pore, these kind of contacts have not statistically been evaluated.

Propensity Scales and Crossvalidation

The propensity scales for the two data sets were calculated as follows: the propensity of an amino acid for a certain contact type (e.g., buried) is calculated by dividing the fraction of this type of contact showing the specified amino acid by the fraction of all other contacts showing the specified amino acid and by taking the logarithm of this value. The propensities of all amino acids at different positions relative to the contact are stored in a 2D matrix (see link at: <http://bioinformatics.charite.de/>).⁴² Hence, when the contact type for a certain amino acid position is

calculated, all the neighboring positions up to a specified window width are assessed as well.

The predictions are carried out on the amino acid sequences of the transmembrane helices that are defined in the corresponding PDB files. The outcome of the predictions is validated by a helix- or protein-wise crossvalidation. Because no redundant information is applied, the different propensity scales can be directly used to evaluate their suitability for the prediction of the two data sets. The mean values given in the results section were calculated on the basis of helix-wise crossvalidations. Helices to be predicted are scored by propensities determined from the rest of the data and are sorted by their scores. Scores of peptides in the test set were calculated by multiplying the positional specific propensities of amino acids in different windows around the central residue. For each score threshold, sensitivity (percentage of correctly predicted contact residues) and specificity (percentage of correctly predicted noncontact residues) can be calculated. The area under the curve (AUC) obtained by plotting sensitivity versus specificity is known as AUC value and is a measure of the quality of the prediction method.⁴³

The quality of the prediction is dependent on the width of the window considered for the prediction. Nevertheless, this approach is limited by the fact that the total amount of data decreases, when it is increasingly only the central amino acid positions of the transmembrane helix that are taken into account. But, the terminal positions of the transmembrane helices are influenced by the polar lipid head groups and the aqueous milieu and bear distinct structural signatures.^{22,25} Using a window of ± 7 , an average of two helix turns is abolished at both termini of the transmembrane helices, what approximates to the mean value plus the standard deviation σ (1.2 ± 1 turn) expected to protrude into the polar milieu.²⁵ Because the expansion of the lipid tail spanning helical segments are not determined by secondary structure prediction tools,⁴⁴ the window of ± 7 was taken for the prediction of contact types. Accordingly, helices shorter than four windings are not gathered by our method (Table I). The link for the program to predict helix-helix contacts will be found at <http://bioinformatics.charite.de/>.

RESULTS AND DISCUSSION

Sequential Motifs of Helix-Helix Interaction in Channels and Membrane Coils

The common characteristics of the amino acid compositions of channels and membrane coils, as described in our previous analysis,^{15,25} are still valid for the updated set of proteins and are not further discussed here. When the propensity scales for helix-helix contacts of the two data sets are weighted against each other, there is a striking difference (Fig. 1): Compared to membrane coils, there is no preference of polar amino acids for helix-helix contacts in channels. Because these residues are even two times more frequent in channels as in membrane coils,^{15,25} it seems obvious that polar amino acids are important for the function of channels. However, wide open pores exposing large helical patches to the polar milieu are only found in a

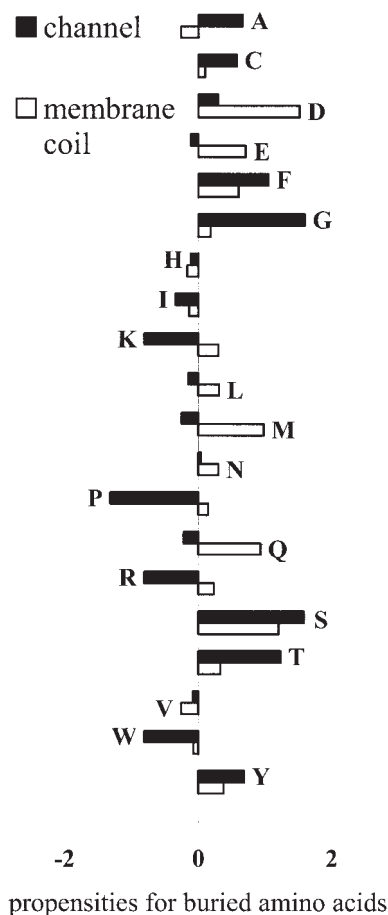


Fig. 1. Propensities for buried amino acids of channels ($n = 665$) and membrane coils ($n = 1190$), excluding at an average two turns at the termini of the transmembrane helices.

subset of the investigated channel structures. In fact, there is only a slight enrichment of polar and charged residues in pore-lining compared with buried positions.²² Because polar and charged residues are commonly not exposed to the hydrocarbon core region of the membrane, the polar residues in channels must be preferred at other locations in the structure.

The analysis of molecular packing densities provides a first step towards the solution for this puzzle. Compared to membrane coils, the side chains of polar amino acids are typically more loosely packed in channels and are twice as often found at protein cavities.¹⁵ Practically all of these cavities that are lined along the pores of the channel are large enough to carry bulk solvent. We therefore subsume that at least some of these cavities are filled with bulk solvent, which is often not recorded in the PDB structures.¹⁵ Bulk solvent may already play a crucial role during the first steps of the folding of channels. It has been demonstrated that it is bound to the polar residues of the S4 helix of the KvAP K^+ channel, when the helix is inserted into the membrane.¹⁶ The ability of polar residues to bind bulk solvent may accordingly be essential for the folding and function of channels. In contrast to membrane

coils, it is therefore likely that polar residues do not promote strongly for helix-helix interactions in channels.

Motifs with small amino acids drive the assembly of transmembrane helices.^{11,45} These motifs are characteristically found in all helical membrane proteins,^{28,30,33} giving rise to the proposal that transmembrane helices are packed more tightly as helices of globular proteins.²⁹ This tight packing was supposed to compensate for the lack of the hydrophobic effect as a driving force for helix-helix interaction in membranes. But recent analysis offer another point of view, showing that the smaller distance of closest approach of two helices does not necessarily correlate with higher packing densities.^{15,31} This is particularly evident for helix-helix interactions of channels which packing densities are low, although the distances between the helix backbones are closer compared to membrane coils.¹⁵ In channels, accordingly, there is a clear preference for small and weakly polar amino acids to be buried between helices. This, however, is not observed for membrane coils (Fig. 1).

The interaction between two helices is labeled by multiple contacts. These are subsumed as sequential patterns, which is the alternate preference of certain residues within the primary structure.^{18,46,47} The main theme observed in helical membrane proteins is patterns of the small residues Gly, Ala, and Ser.²⁸ Small amino acids at *i* and *i* + 4 positions promote the assembly of transmembrane helices, while other residues put at these positions disrupt it.^{11,45} To detect sequential patterns that are characteristic for channels or membrane coils, the correlation between the propensities of neighboring amino acids in the 2D matrices was graphed. Accordingly, residues that promote or prevent helix-helix interactions, cluster every fourth residue in channels, while in membrane coils the typical periodicity of helix-helix contacts is 3.5 [Fig. 2(a)].

These characteristics highlight a basic difference between channels and membrane coils: In channels, about two-third of the helices cross at right-handed angles, where the contacts are typically achieved every fourth residues.^{7,25} In membrane coils, however, about two-thirds of the helices cross at left-handed angles and the helices characteristically make contacts every 3.5 residues. But the optimal geometry of helix-helix interaction that allows for the formation of multiple C α -H---O bonds is only given in right-handed helix-helix motifs.³³ Because these bonds are of considerable strength inside the membrane, the affinity of left-handed Gly heptad repeat motifs to drive membrane helix association is likely to be less strong than in right-handed motifs.⁴⁵ Therefore, energetically considerations offer a good explanation for the fact that there is a clear preference of polar residues to be found in helix-helix contacts of membrane coils, but not in channels. In the Photosystem I the relationship of C α -H---O bonds to conventional inter helical hydrogen bonds is 1.5.⁴⁸ By contrast, the interactions of transmembrane helices of the KcsA potassium channel or the MscL mechanosensitive channel are not stabilized by conventional inter helical hydrogen bonds at all (not shown). In view of

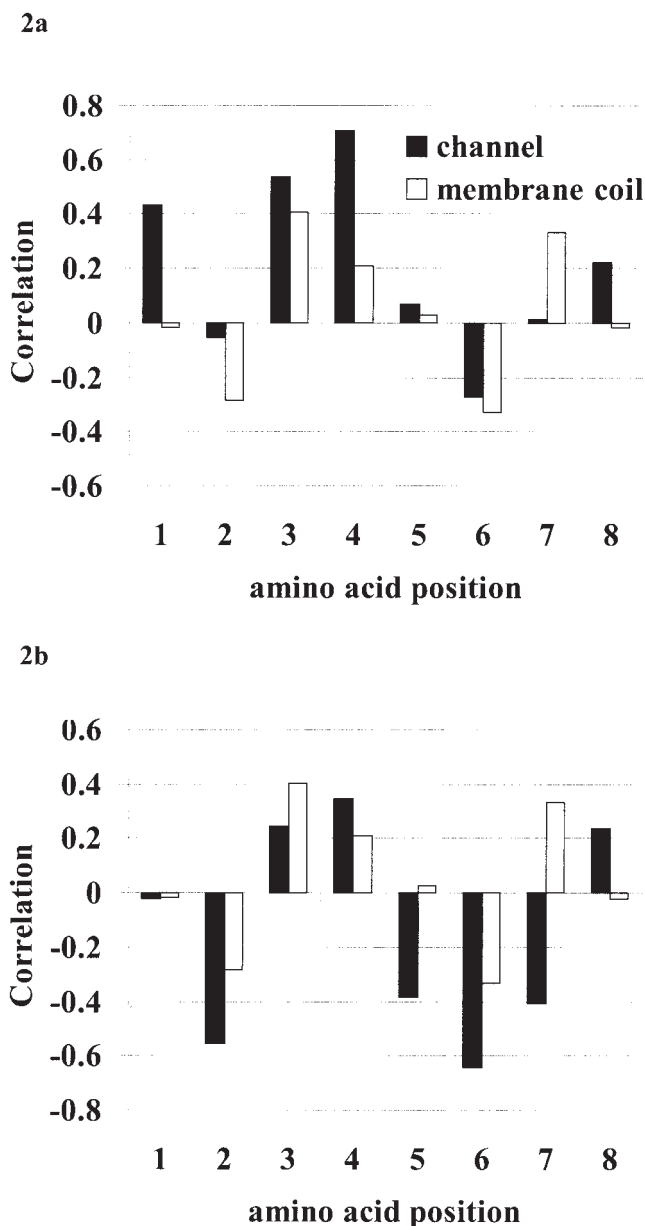


Fig. 2. Correlation between amino acid propensities of helix-helix contacts and the sequence positions adjoining to it, of (a) all residues and (b) the subset of polar residues (Asn, His, Pro, Gln, Arg, Glu, Lys, Asp)¹⁰ in channels and membrane coils. The periodicity of contacts deduced from these correlations for all residues is four (3, 4, and 8) in channels and 3.5 (3, 4, and 7) in membrane coils. In contrast to membrane coils, polar amino acids are not preferred at helix-helix contacts of channels (see Fig. 1). Therefore, the negative correlations of polar amino acids at positions 2 and 6 signify that these positions may be preferred by polar amino acids in channels, but not in membrane coils.

the fact that interhelical hydrogen bonds drive strong interactions in membrane proteins,⁴⁹ we suggest that polar interactions are essential for the folding, stability, and rigidity of membrane coils.

The sequence combining the highest propensity values of amino acids for the helix-helix interactions of membrane coils is (D > Q > G > S)_{-4 · -3 · -2 · -1} (D > S > M > Q)_{0 · +1 · +2 ·} (N > D > S > G)₊₃, with the subscript letters

labeling the propensities at the contact₀, the preceding_{-1 to -4} and subsequent positions_{+1 to +3}. The sequence constructed to ideally accomplish the nature of helix–helix contacts in channels is (F > G > S > T)₋₄ · ₋₃ · ₋₂ · ₋₁ (G > S > T > F)₀ · ₊₁ · ₊₂ · ₊₃ · (T > G > A > S)₊₄. In contrast to channels, Gly is preferred at every third and fourth neighboring position, but not at the helix–helix contact itself in membrane coils, reflecting the characteristics of the knobs into holes packing of left-handed interactions.^{45,46} Instead, residues with polar side chains and Met are preferably buried at the helix–helix contacts of membrane coils. Only Ser is preferred at helix–helix contacts of both data sets, manifesting that motifs of Ser are strong determinants of stability in left- and right-handed interaction motifs.^{47,50}

Therefore, as judged from the amino acid propensities and from the periodicity of helix–helix contacts, the mode of helix–helix interactions is clearly different in the two data sets: in channels the assembly of helices is achieved by the burial of small and weakly polar amino acids, which is followed by the formation of multiple Cα–H···O bonds. In membrane coils, however, the stability of helix–helix interactions is mainly accomplished by polar amino acids, which are placed inside heptad repeat motifs of small amino acids.

It has been proposed that the detailed geometrical fit between two associating helices determines the specificity of helix–helix contacts, whereas the stability is manifested by the formation of hydrogen bonds.¹² We suggest that the observed patterns of small amino acids essentially determine the specificity of the helix–helix assembly in all helical membrane proteins. The stability, however, is realized another way when membrane proteins with different functions are compared. In membrane coils, helix–helix contacts are frequently stabilized by polar interactions, which also results in higher molecular packing densities.¹⁵ In channels, polar residues are preferred at sequence positions ± 2 and ± 6 relative to the helix–helix contact [Fig. 2(b)]. If these positions contact the polar milieu during the folding, the interfaces of neighboring transmembrane helices could be orientated towards each other, promoting the folding of the channel.

Comparison of the Propensity Scales for Helix–Helix Contacts with Other Scales

The specific sequence-coding for transmembrane helices to be identified by the endoplasmic reticulum Sec61 translocon has recently been determined, presenting the first “biological” hydrophobicity scale.¹⁰ Considering this scale, it is likely that the common weakly polar amino acids are buried in helix–helix contacts: these residues are probably too polar for the lipophilic surrounding,²³ but at the same time not hydrophilic enough for the polar milieu.^{15,16} This would signify that the folding of channels is ruled by the preference of subsets of hydrophobic, weakly polar, and polar residues to contact the hydrophobic or the polar milieu. In fact, nonhydrophobic residues are preferred at helix–helix contacts up to a certain threshold. This threshold is marked by a big step within the biological hydrophobicity scale at His and Pro where the polar amino acids

start,¹⁰ which again tend to interact with the polar milieu. Likewise, there is a specific role discernible for the aromatic residues and especially for Trp to intermingle with the polar lipid head groups.^{23,24}

The positions of polar and aromatic residues within the helix specify, whether the helix inserts into the membrane or not.²⁰ As a result, the propensities of these residues for helix–helix contacts in the membrane do not correlate with the biological hydrophobicity scale. By contrast, there is a remarkable correlation for the subset of all nonpolar and nonaromatic residues, when the channels scale for helix–helix contacts is compared with the biological hydrophobicity scale [Fig. 3(a)]. Moreover, there is a good correlation between the propensities of the hydrophobic and weakly polar amino acids to contact the residue of another helix or the membrane [Fig. 3(b)]. The same is valid when other propensity scales for helix–membrane contacts, derived from sequential (kPROT) or structural data are taken instead (kPROT:³⁷ $R = 0.85$, STRUCTURE SP:³⁹ $R = 0.92$, TMLIP1-C:²³ $R = 0.93$).

However, when the entire scales are compared, there is no significant similarity between the channels scale for helix–helix contacts and these propensity scales for helix–membrane contacts. This accounts for the fact that the properties of the polar and aromatic residues to contact the helix or the membrane are modulated by their tendency to get in touch with the lipid head group region or bulk solvent.^{23,24} Finally, there is a moderate correlation, when the propensities of the weakly polar and hydrophobic residues for helix–helix contacts of the channel scale is compared with another helix–helix packing scale (Packing scale 2:³² $R = 0.88$). The reason why these two scales are fairly different is caused by the fact that they have been calculated with distinct methods, as discussed previously.¹⁵

For the excellent correlation of the channels scale for helix–helix contacts with the biological hydrophobicity scale, we conclude that the solubility of weakly polar and hydrophobic residues is the main determinant for the assembly of the transmembrane helices of channels. As a consequence, it seems that the information, which is substantial for the folding of the tertiary structure of channels, is already recognized by the endoplasmic reticulum translocon. The recent model to describe the folding of helical membrane proteins reproduces that the tertiary structure is at least partially completed before a membrane protein is leaving the translocon.¹³ A current study on the *in vivo* folding of the aquaporin-4 water channel actually manifests that the helices move out of the Sec-translocon in succession but may stay nearby for an extended period of time.¹⁹ Therefore, the above-mentioned conclusion is also supported by recent *in vivo* folding models.

Nevertheless, it seems much more complicated to elucidate the driving forces that are responsible for the assembly of helices in membrane coils. It has been mentioned above that the polar residues promote for the packing of transmembrane helices. But, as reported for Rhodopsin, the binding of coenzymes and ligands is additionally

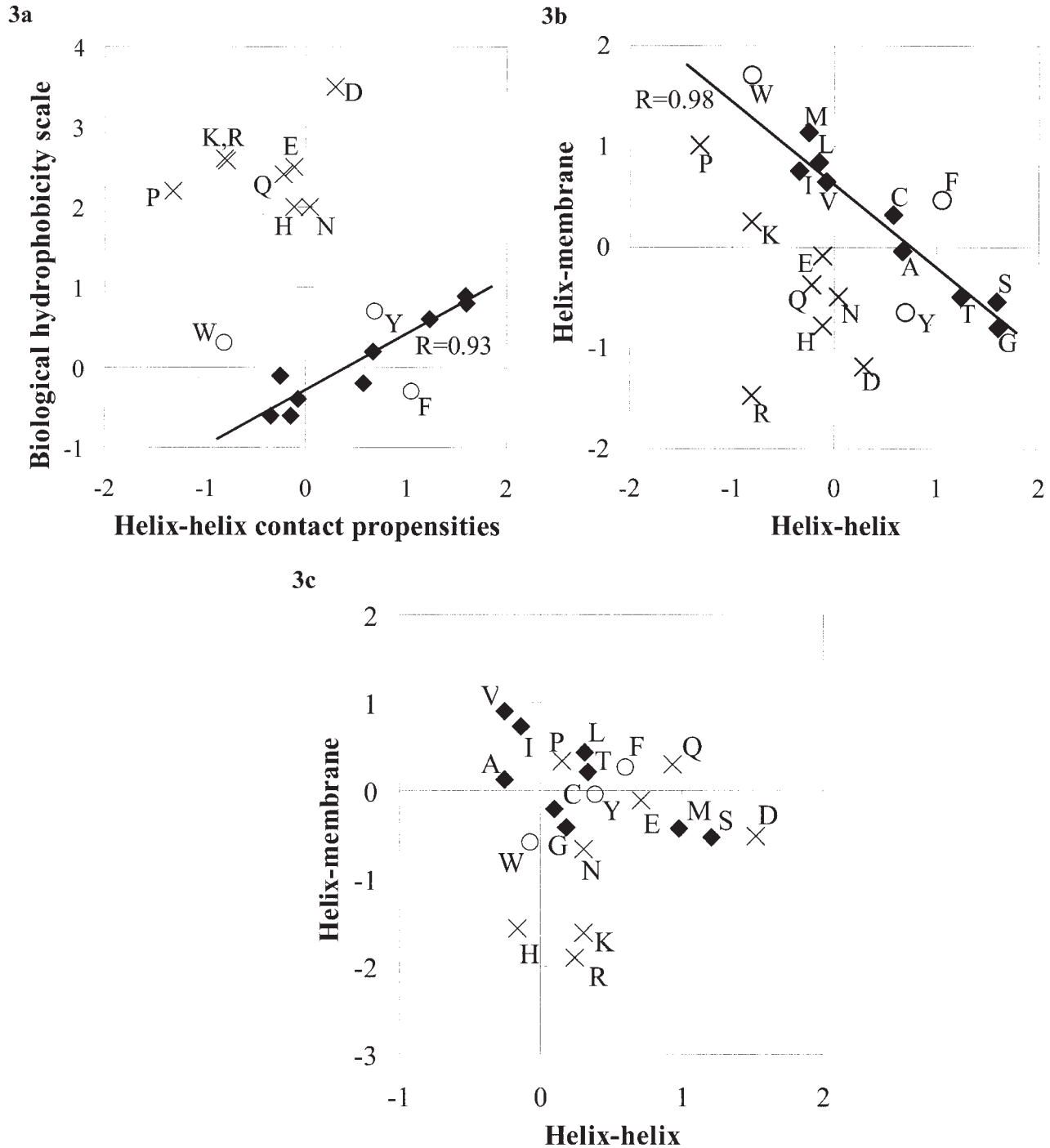


Fig. 3. Correlation between the amino acid propensities of (a) buried residues in transmembrane helices of channels and the biological hydrophobicity scale,¹⁰ of buried and exposed residues in (b) channels and (c) membrane coils. Aromatic residues are depicted as small circles, polar residues as crosses, and weakly polar and hydrophobic residues as small filled diamonds.

stabilizing the 3D structure of many membrane coils.⁵¹ This mechanism probably also accounts for the fact that there is no significant correlation between the propensities of exposed and buried residues in membrane coils [Fig. 3(c)]. Besides, there is no correlation between any other scale mentioned here, and the preference of a residue to be buried at helix-helix contacts of mem-

brane coils (not shown). We therefore expect that the binding of coenzymes and ligands is essential for the proper assembly of the transmembrane helices in most membrane coils. These interactions might compensate for the lack of contacts with bulk solvent, which we suppose to be one of the driving forces for the folding of channels.

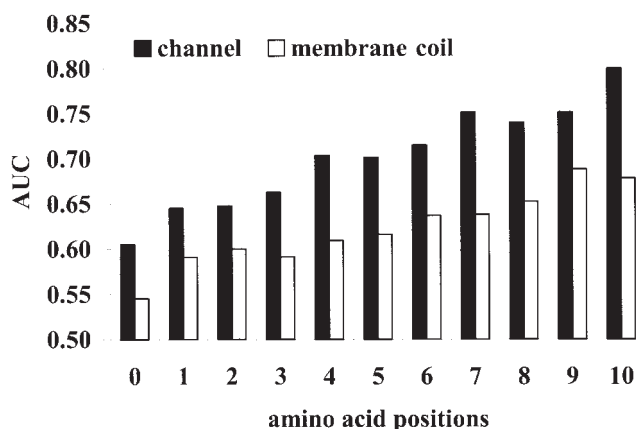


Fig. 4. AUC-values for the prediction of helix-helix contacts of transmembrane helices in channels and membrane coils. The prediction quality rises, when the preceding and subsequent amino acid positions are taken into account.

The Prediction of Helix-Helix Interactions in Channels and Membrane Coils

All the following predictions are the results of crossvalidations. This means that a predicted helix or protein structure is not part of the learning data set, which composes of all other helices or structures of the channels or membrane coils data set. Therefore, it is a significant finding that the quality of prediction for helix-helix contacts is improved when the composition of the neighboring amino acids is also taken into account (Fig. 4). Moreover, best predictions are achieved when the appropriate scale is used (Table I), indicating that the investigated channel structures share structural characteristics that are not represented well by the membrane coils matrix and vice versa.

The matrix prediction method is accordingly sensitive for the detection of sequential patterns, which are highlighted in Figure 5. This tendency is generally more noticeable for channels, where the AUC values rise periodically when the positions ± 4 , ± 7 , and ± 10 are additionally considered (Fig. 4). For the window of ± 7 , which is the standard of all our predictions (see Methods) the AUC value for the prediction of helix-helix contact is 0.76. This is a significant enhancement of +16% compared to the contact prediction, where the neighboring amino acid composition is not taken into account. By contrast, the prediction of helix-helix contacts is lower for membrane coils, yielding only an AUC value of 0.64.

Accordingly, the recognition of sequential motifs, which is the base of our method to predict helix-helix contacts is less efficient for membrane coils as that for channels. This difference may have several reasons. The knobs into hole packing is the prevalent structural motif in left-handed interactions of membrane coils.⁴⁶ Residues with bulky side chains as Met, Leu, and Phe pack into the grooves formed by small amino acids. But Leu and Phe show no clear preference to be buried or exposed [Fig. 3(b)]. Small amino acids are not really preferred at helix-helix contacts and the polar amino acids, which are generally buried in

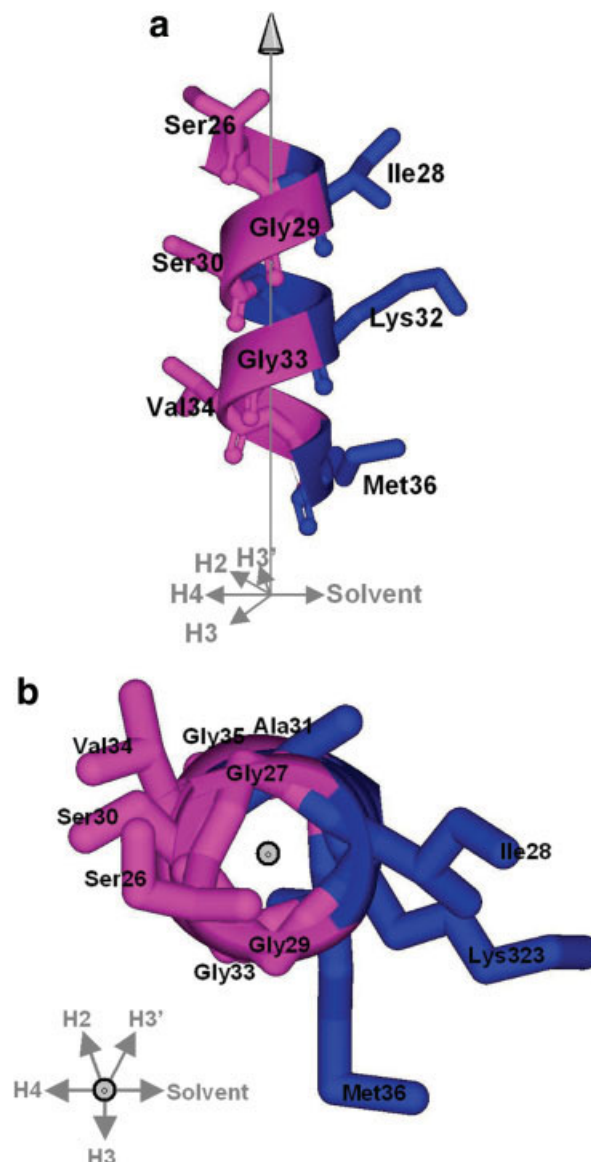


Fig. 5. (a) The contacts of helix 1 of the membrane rotor of the V-type ATPase are predicted accurately (PDP-code: 2bl2, side view). The contacts to the four helix neighbors (in purple, H2-4 is helix 2 to 4 of the same, H3' of the other subunit) are established by small and weakly polar residues, four positions apart in the sequence. In contrast, residues with bulky or polar side chains are in contact with the bulk solvent and the lipids (in blue), located between the membrane rotor ring and the central stalk⁵² (not depicted). (b) Top view of Figure 5(a).

membrane coils, are not very frequent in transmembrane helices. Finally, it is not considered by our method that the stability of membrane coils is additionally enhanced by the incorporation of ligands and coenzymes, which interact with all different kinds of residues.⁵¹ The AUC values for the prediction of helix-membrane contacts, however, are similar for membrane coils (0.72) and channels (0.68) (not shown).

When the outcome of our predictions are compared to another approach where the AUC values for a subset of 11 membrane proteins were calculated with the help of single

amino acid propensities, our method performs better at an average of about 10% for channels and 7% for membrane coils (Table I).³⁹ Moreover, it is shown that the incorporation of conservation criteria enhances the prediction of those membrane proteins where homologues proteins are available by another 10%.^{38,39} However, for many membrane proteins, appropriate sequence alignments do not exist and the prediction of helix contact types is based on the usage of propensity scales. Besides, without considering helix-helix propensities it is not clear whether a conserved residue is essential for the function or for the stability of a protein, which additionally complicates the modeling process.²²

Our methods to predict helix-helix or helix-membrane contacts provide independent information about the probability of a certain residue to be buried or exposed. To further refine our predictions of helix-helix contacts, the two scores have been combined. If only those amino acids were assigned as helix-helix contacts, which prediction score is above a certain threshold (two times larger as the score for the helix-membrane prediction), 56% of all helix-helix contacts are predicted with 79% correctness (not shown). For membrane coils, nearly 70% of all helix-helix contacts are predicted with an accuracy of 73% (not shown).

The advantage of the latter approach is that especially those residues are predicted at high accuracy that are placed in typical sequential patterns (see Fig. 5). During the completion of this manuscript, an analysis was published carrying experimental evidence that the "glycine zipper motif," which is actually one of the main motifs in right-handed transmembrane helix interactions, is also important in a wide range of normal and pathological cellular processes.¹⁸ Here, we present a prediction method that is highly sensitive in the detection of these and other structural motifs that are supposed to be strong determinants of the folding of transmembrane helices in channels.

CONCLUSION

We calculated propensity scales for 27 different high-resolution structures of helical membrane proteins subdivided into membrane coils (surface receptors and redox proteins) and channels (membrane channels and transporters). In membrane coils, the polar amino acids are strongly favored in helix-helix contacts. We conclude that these residues are essential for the folding, stability, and rigidity of membrane coils. By contrast, polar amino acids are supposed to have a rather small impact on the assembly of the transmembrane helices of channels. These residues are found at sequential positions that are preferably not buried, but exposed to the polar milieu during the folding. The helix assembly however is mainly accomplished by weakly polar residues that cluster every fourth residue. These proteins are fairly stabilized by the onset of multiple C α -H \cdots O bonds, that are typically found in helices crossing at right-handed angles, which is the prevalent motif of channels.

The propensities of the subset of hydrophobic and weakly polar amino acids to be buried or exposed strongly correlates with the now-available biological hydrophobicity scale. Therefore, our analysis confirms that the folding of channels can be explained by a simple rule where the weakly polar residues pack together, while polar residues and hydrophobic residues are presupposed to contact the aqueous and lipophilic milieu, respectively. The prediction of helix contact types is improved significantly when the structural patterns of membrane proteins with different functions are considered. More than half of all helix-helix contacts of channels are predicted with 79% accuracy and nearly three-quarters of the helix-helix contacts in membrane coils are predicted with the precision of 73%. The major challenge of our method is that no homology is needed to predict reliably those helix-helix contacts that show the typical sequence patterns of channels or membrane coils.

ACKNOWLEDGMENTS

We would like to thank Cornelius Frömmel and Kristian Rother for helpful discussions.

REFERENCES

1. Jiang Y, Lee A, Chen J, Cadene M, Chait BT, MacKinnon R. The open pore conformation of potassium channels. *Nature* 2002;417:523–526.
2. Toyoshima C, Nomura H, Tsuda T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature* 2004;432:361–368.
3. Rosenbusch JP. Stability of membrane proteins: relevance for the selection of appropriate methods for high-resolution structure determinations. *J Struct Biol* 2001;136:144–157.
4. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
5. Montelione GT, Anderson S. Structural genomics: keystone for a Human Proteome Project. *Nat Struct Biol* 1999;6:11–12.
6. White SH, Wimley WC. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct* 1999;28:319–365.
7. Popot JL, Engelman DM. Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* 2000;69:881–922.
8. Lehnert U, Xia Y, Royce T, Goh C, Liu Y, Senes A, Yu H, Zhang Z, Engelman D, Gerstein M. Computational analysis of membrane proteins: genomic occurrence, structure prediction and helix interactions. *Q Rev Biophys* 2005;37:1–26.
9. White SH, von Heijne G. The machinery of membrane protein assembly. *Curr Opin Struct Biol* 2004;14:397–404.
10. Hessa T, Kim H, Bihlmaier K, Lundin C, Boekel J, Andersson H, Nilsson I, White SH, von Heijne G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 2005;433:377–381.
11. Lemmon MA, Engelman DM. Specificity and promiscuity in membrane helix interactions. *FEBS Lett* 1994;346:17–20.
12. Schneider D. Rendezvous in a membrane: close packing, hydrogen bonding, and the formation of transmembrane helix oligomers. *FEBS Lett* 2004;577:5–8.
13. Stevens TJ, Mizuguchi K, Arkin IT. Distinct protein interfaces in transmembrane domains suggest an in vivo folding model. *Protein Sci* 2004;13:3028–3037.
14. Deusch C. Potassium channel ontogeny. *Annu Rev Physiol* 2002;64:19–46.
15. Hildebrand PW, Rother K, Goede A, Preissner R, Frommel C. Molecular packing and packing defects in helical membrane proteins. *Biophys J* 2005;88:1970–1977.
16. Freitas JA, Tobias DJ, von Heijne G, White SH. Interface connections of a transmembrane voltage sensor. *Proc Natl Acad Sci USA* 2005;102(42):15059–15064.

17. Hristova K, White SH. An experiment-based algorithm for predicting the partitioning of unfolded peptides into phosphatidylcholine bilayer interfaces. *Biochemistry* 2005;44:12614–12619.
18. Kim S, Jeon TJ, Oberai A, Yang D, Schmidt JJ, Bowie JU. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc Natl Acad Sci USA* 2005;102(40):14278–14283.
19. Sadlish H, Pitonzo D, Johnson AE, Skach WR. Sequential triage of transmembrane segments by Sec61alpha during biogenesis of a native multispansing membrane protein. *Nat Struct Mol Biol* 2005;12:870–878.
20. Hessa T, White SH, von Heijne G. Membrane insertion of a potassium-channel voltage sensor. *Science* 2005;307:1427.
21. Bowie JU. Helix-bundle membrane protein fold templates. *Protein Sci* 1999;8:2711–2719.
22. Eyre TA, Partridge L, Thornton JM. Computational analysis of (alpha)-helical membrane protein structure: implications for the prediction of 3D structural models. *Protein Eng Des Sel* 2004;17:613–624.
23. Adamian L, Nanda V, DeGrado WF, Liang J. Empirical lipid propensities of amino acid residues in multispans alpha helical membrane proteins. *Proteins* 2005;59:496–509.
24. Ozdirekcan S, Rijkers DT, Liskamp RM, Killian JA. Influence of flanking residues on tilt and rotation angles of transmembrane peptides in lipid bilayers. A solid-state ²H NMR study. *Biochemistry* 2005;44:1004–1012.
25. Hildebrand PW, Preissner R, Frömmel C. Structural features of transmembrane helices. *FEBS Lett* 2004;559:145–151.
26. Popot JL, Engelman DM. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* 1990;29:4031–4037.
27. Jones DT, Taylor WR, Thornton JM. A mutation data matrix for transmembrane proteins. *FEBS Lett* 1994;339:269–275.
28. 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.
29. 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.
30. Adamian L, Liang J. Helix–helix packing and interfacial pairwise interactions of residues in membrane proteins. *J Mol Biol* 2001;311:891–907.
31. Gimpelev M, Forrest LR, Murray D, Honig B. Helical packing patterns in membrane and soluble proteins. *Biophys J* 2004;87:4075–4086.
32. Liu W, Eilers M, Patel AB, Smith SO. Helix packing moments reveal diversity and conservation in membrane protein structure. *J Mol Biol* 2004;337:713–729.
33. Senes A, Ubarretxena-Belandia I, Engelman DM. The Calpha–H...O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci USA* 2001;98:9056–9061.
34. Fleishman SJ, Ben-Tal N. A novel scoring function for predicting the conformations of tightly packed pairs of transmembrane alpha-helices. *J Mol Biol* 2002;321:363–378.
35. Park Y, Elsner M, Staritzbichler R, Helms V. Novel scoring function for modeling structures of oligomers of transmembrane alpha-helices. *Proteins* 2004;57:577–585.
36. Taylor WR, Jones DT, Green NM. A method for alpha-helical integral membrane protein fold prediction. *Proteins* 1994;18:281–294.
37. 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.
38. Fleishman SJ, Harrington S, Friesner RA, Honig B, Ben-Tal N. An automatic method for predicting transmembrane protein structures using cryo-EM and evolutionary data. *Biophys J* 2004;87:3448–3459.
39. Beuming T, Weinstein H. A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. *Bioinformatics* 2004;20(12):1822–1835.
40. Preissner R, Goede A, Frommel C. Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. *J Mol Biol* 1998;280:535–550.
41. Rother K, Preissner R, Goede A, Frömmel C. Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins. *Bioinformatics* 2003;19:2112–2121.
42. Lorenzen S, Peters B, Goede A, Preissner R, Frommel C. Conservation of cis prolyl bonds in proteins during evolution. *Proteins* 2005;58:589–595.
43. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–577.
44. Tusnady GE, Simon I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* 2001;17:849–850.
45. Lear JD, Stouffer A, Gratkowski H, Nanda V, DeGrado WF. Association of a model transmembrane peptide containing Gly in a heptad sequence motif. *Biophys J* 2004.
46. Langosch D, Heringa J. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins* 1998;31:150–159.
47. Adamian L, Liang J. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* 2002;47:209–218.
48. Loll B, Raszewski G, Saenger W, Biesiadka J. Functional role of C(alpha)-H...O hydrogen bonds between transmembrane alpha-helices in photosystem I. *J Mol Biol* 2003;328:737–747.
49. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM. Interhelical hydrogen bonding drives strong interactions in membrane proteins [see comments]. *Nat Struct Biol* 2000;7:154–160.
50. Schneider D, Engelman DM. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. *J Mol Biol* 2004;343:799–804.
51. Okada T, Ernst OP, Palczewski K, Hofmann KP. Activation of rhodopsin: new insights from structural and biochemical studies. *Trends Biochem Sci* 2001;26:318–324.
52. Murata T, Yamato I, Kakinuma Y, Leslie AG, Walker JE. Structure of the rotor of the V-Type Na⁺-ATPase from *Enterococcus hirae*. *Science* 2005;308:654–659.