Interhelical Hydrogen Bonds and Spatial Motifs in Membrane Proteins: Polar Clamps and Serine Zippers

Larisa Adamian and Jie Liang*
Department of Bioengineering, the University of Illinois at Chicago, Chicago, Illinois

ABSTRACT **Polar and ionizable amino acid residues are frequently found in the transmembrane (TM) regions of membrane proteins. In this study, we show that they help to form extensive hydrogen bond connections between TM helices. We find that almost all TM helices have interhelical hydrogen bonding. In addition, we find that a pair of contacting TM helices is packed tighter when there are interhelical hydrogen bonds between them. We further describe several spatial motifs in the TM regions, including "Polar Clamp" and "Serine Zipper," where conserved Ser residues coincide with tightly packed locations in the TM region. With the examples of halorhodopsin, calcium-transporting ATPase, and bovine cytochrome c oxidase, we discuss the roles of hydrogen bonds in stabilizing helical bundles in polytopic membrane proteins and in protein functions. Proteins 2002;47:209–218.**

Key words: membrane protein; helical packing; hydrogen bonds; alpha shape; spatial motifs

INTRODUCTION

Polar and ionizable amino acids constitute 20–22% of all residues in the transmembrane (TM) helices, as revealed by statistical analysis of the composition of predicted transmembrane α-helices of polytopic membrane proteins.[1,2] Among these, amino acid residues S, T, and C are the most frequently found polar residues, accounting for a combined total of 13% of all residues in the predicted TM regions.[1] However, these residues have unfavorable transfer free energy from water to phospholipid bilayer, as experimentally determined in model systems.[3] Although polar and ionizable amino acids are often key residues in functional sites of membrane proteins,[4–8] only a fraction of them is directly involved in protein function. What are the roles of polar amino acid residues in the TM region?

Polar groups of amino acids in the TM helices are often involved in networks of H-bonds, which are found in high-resolution structures of bacteriorhodopsin[4] and rhodopsin.[7] Recent experimental data[9,10] suggest that interhelical H-bonds may be important for stability and specificity of helical association in the TM regions of membrane proteins. In these experiments, the introduction of interhelical H-bond between two engineered transmembrane helices derived from GCN4 leucine zipper[9,10] resulted in strong interhelical association. The dimer association of engineered leucine zipper is found to be more stable than the helical association driven by van der Waals interactions and weak C_α-H—O hydrogen bond, as seen in glycophorin A.[11] In addition, the propensity for engineered peptides to form stable oligomers correlates with the chemical nature of the incorporated amino acid residue.[12] Peptides containing amino acids with two polar atoms (N, Q, D, E) form stable trimeric structures, whereas peptides containing amino acids with less than two polar atoms have a much weaker tendency to associate.

The energetics of H-bonds is largely electrostatic in nature and is dominated by the Coulombic interaction between the partial effective charges on the donor and the acceptor atoms. This interaction depends on the effective dielectric constant of the environment and the distance separating the partial charges. Consequently, the formation of H-bonds in phospholipid bilayer is likely to be far more energetically favorable because of the low effective dielectric environment of the lipids. This has been demonstrated experimentally by the measurement of the equilibrium of forming intramolecular hydrogen bond ($\mathbf{K^{HB}}$) in a series of substituted salicylate monoanions as a function of $\mathbf{\Delta pK_a}$ in DMSO and water.[13]

The increasing number of three-dimensional structures of membrane proteins[4–8,14–16] allows detailed structural analysis of packing[17,18] and interhelical contacts[19] in the TM regions. A recent study found that polar-polar atomic interactions constitute about 4% of all atomic interhelical contacts in membrane and in soluble proteins.[19] However, the patterns of polar-polar contacts are different. In soluble proteins, only salt bridge residue pairs (D-R, D-K, E-R) have a high propensity for interhelical polar-polar atomic contacts, whereas a greater variety of high propensity residue pairs in interhelical contact is found in membrane proteins. In addition to salt bridges, there are residue pairs between ionizable and polar residues (D-Y, Y-R), as well as residue pairs between polar nonionizable residues (Q-S, S-S), all with high interhelical interaction propensity. Other ionizable-polar residue pairs (E-N, H-T, H-Q, K-Q, K-N, N-R, Q-R, S-R, S-D, D-N) as well as polar

Grant sponsor: NSF; Grant number: DBI-0078270, MCB998008; Grant sponsor: ACS (Petroleum Research Fund); Grant number: 35616-G7.

*Correspondence to: Department of Bioengineering, the University of Illinois at Chicago, 851 S. Morgan Street, Rm. 218, MC-063, Chicago, IL 60607. E-mail: jliang@uic.edu

Received 21 August 2001; Accepted 12 November 2001

TABLE I. Set of Membrane Proteins Used in This Study

PDB ID	Protein	Res, Å	Ref.
1ar1	Cytochrome c oxidase (*Paracoccus denitrificans*)	2.7	Ostermeier et al.[20]
1ehk	Cytochrome c oxidase (*Thermus thermophilus*)	2.4	Soulimane et al.[14]
1Ocr	Cytochrome c oxidase (*Bos taurus*)	2.4	Yoshikawa et al.[21]
1be3	Cytochrome bc1 complex (*Bos taurus*)	3.0	Iwata et al.[42]
1dxr	Photosynthetic Reaction Center (*Rhodopseudomonas viridis*)	2.0	Lancaster et al.[16]
1c3w	Bacteriorhodopsin (*Halobacterium salinarum*)	1.6	Luecke et al.[4]
1e12	Halorhodopsin (*Halobacterium salinarum*)	1.8	Kolbe et al.[6]
1f88	Rhodopsin (*Bos taurus*)	2.8	Palczewski et al.[7]
1fum	Fumarate Reductase Flavoprotein Subunit (*E. coli*)	3.3	Iverson et al.[15]
1qla	Fumarate Reductase Flavoprotein (*Wolinella succinogenes*)	2.2	Lancaster et al.[16]
1fx8	Glycerol-Conducting Channel (*E. coli*)	2.2	Fu et al.[5]
1bl8	Potassium Channel Protein (*Streptomyces lividans*)	3.2	Doyle et al.[43]
1eul	Calcium-Transporting ATPase (*Oryctolagus cuniculus*)	2.6	Toyoshima et al.[8]

residue pairs (Q-T, Q-Y, S-C, N-N, N-Q, N-S, N-Y, N-W) may also have high propensity to form interhelical polar-polar interactions, although the size of the data set is too small to make definite conclusions on these residue pairs.

In this study, we analyze the interhelical H-bonds in a set of 13 TM protein structures. Our goal is to study the distributions of H-bonds between helices in the TM region, the composition of amino acid residues that are involved in H-bonds in TM helices, and the three-dimensional spatial motifs of polar residues. Our results indicate that almost every helix in this data set is connected by at least one H-bond to the closest neighboring helix. We also find that pairs of interacting helices with H-bonds are packed tighter than those without H-bonds. H-bonds between two side chains and H-bonds between a side chain and a backbone carbonyl oxygen are found to occur with approximately the same frequency (51 and 49%, respectively). We also describe two novel spatial motifs identified in this study, namely, polar clamp and serine zipper, and discuss the correlation with their structural packing and sequence variation. We further discuss the likely structural and functional roles of these spatial motifs in calcium transporting ATPase, halorhodopsin, and in cytochrome c oxidases.

METHODS

Membrane Protein Data

The 13 membrane proteins used in this study are listed in Table I. This data set includes three structures of cytochrome c oxidases from *Paracoccus dentrificans*,[20] *Thermus thermophilus*,[14] and *Bos taurus*.[21] The sequence identity between cytochrome c oxidases from prokaryotes is low (~21%), while the sequence identities between subunits I and II of cytochrome c oxidases *Bos taurus* and *Paracoccus dentrificans* are 48 and 36%, respectively. There are 16 common H-bonds between conserved residues in these two proteins in the TM region. We excluded them from the statistical analysis of H-bonds. Two structures of fumarate reductase flavoprotein subunits (PDB ID 1fum[15] and 1qla[16]) with low sequence identity are also included. All loops in the soluble regions are manually removed, leaving only the alpha helices in the TM regions. As a result, each protein is represented by a bundle of TM helices. Altogether there are 134 unique helices in the data set. Here we analyze only interhelical H-bonds formed between amino acid residues and remove all ligands and water molecules. H-bonds mediated by water molecules are found in many proteins and often implicated in the functions of membrane protein. However, they are not equally well resolved in all structures in the data set, and we do not include H-bonding involving water molecules in this analysis.

Computation of Interhelical Contacts and H-Bonds

Using the alpha shape application program interface kindly provided by Prof. Edelsbrunner and colleagues, a program INTERFACE has been implemented to compute interhelical contacting atoms. INTERFACE uses precomputed Delaunay triangulation and alpha shape. The Delaunay triangulation of membrane proteins is computed using the DELCX program,[22,23] and the alpha shapes are computed using the MKALF program.[22,24] Both can be downloaded from the website of NCSA (http://www.ncsa.uiuc.edu). The van der Waals radii of protein atoms are taken from Tsai et al.[25] To account for uncertainty in the precision of atomic coordinates, the van der Waals radii are incremented by 0.5 Å following Singh and Thornton.[26] The advantage of using INTERFACE compared to methods using distance cut-off is that only nearest neighbor atoms in physical contacts are counted.[19] H-bonds are identified by HBPLUS program[27] using default parameters and allowing exchange of the nearly symmetrical side chains of residues H, Q, and N, since nitrogen, oxygen, and carbon atoms are indistinguishable in electron density maps. Potential H-bonds that would be formed if histidine CD2 was actually ND1, CE1 was NE2, and the oxygens and nitrogens in assignment of N and Q residues were the other way around, were counted.

RESULTS

Hydrogen Bonds Are Commonly Observed Between TM Helices

There are 134 unique TM helices in the data set of 13 proteins. We declare that a contacting helical pair is

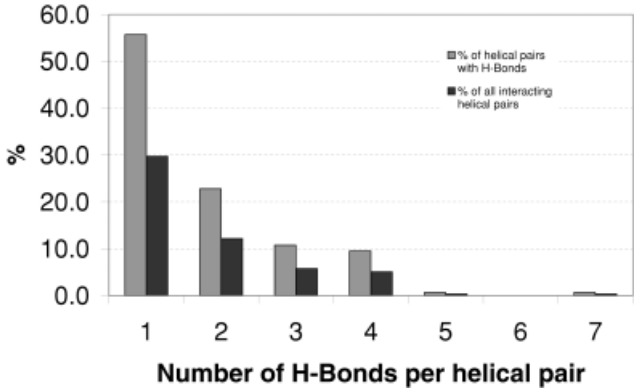

Fig. 1. Distribution of helical pairs connected by hydrogen bonds by the number of H-bonds per contacting helical pair in the full set of helical pairs, which includes helical pairs without H-bonds (black bars) and in the subset of helical pairs that contain H-bonds (gray bars).

formed when two neighboring TM helices have at least one interhelical atomic contact between them. This arbitrary criterion is adopted for convenience. The contact may be an interaction between nonpolar-nonpolar, nonpolar-polar, or polar-polar atoms. INTERFACE program[19] identified 296 such helical pairs in the given set of 13 proteins. The number of interhelical atomic contacts in a helical pair ranges from 1 to 235. On average, the number of contacts per helical pair is about 65. Of all helical pairs, 53% (158) are connected by one or more H-bonds. The distribution of number of H-bonds per pair of interacting helices is shown in Figure 1. Among the subset of helical pairs connected by H-bonds, almost 80% have one or two interhelical H-bonds, although there are cases where three to seven such bonds are observed between a helical pair.

The topological maps of interhelical H-bonds (top view) of the TM region for each of the 13 proteins are shown in Figure 2. Two helices are connected by an edge if there is at least one hydrogen bond between them. The helices are numbered by the order in which they appear in the primary sequence. Figure 2 highlights the presence of extensive H-bond connections between helices and suggests that almost all TM helices are involved in H-bonding. Some of the membrane protein structures are of low resolution (1be3, 3.0Å; 1fum, 3.3Å), which may affect the number of detected H-bonds. The detection of H-bonds depends on the resolution of the structure, as shown by McDonald and Thornton.[27] Here we also find that there are fewer helices connected by H-bonds in low-resolution structures. For example, helix 5 from bovine cytochrome bc1 complex (1be3, 3.0Å) and helix 3 from *Escherichia coli* fumarate reductase flavoprotein (1fum, 3.3Å) do not form H-bonds with any other helix. The additional helix without H-bonding to any other helices is found in high resolution (2.2Å) structure of glycerol conducting channel (1fx8). This helix (8 on Fig. 2) has been found recently to form weak C_α-H—O hydrogen bonds with helix 5.[11] These two helices are tightly packed in a manner reminiscent of glycophorin A.[11,28]

Contacting Helices With H-Bonds Are Packed Tighter

We find that the distributions of the number of total interhelical atomic contacts between neighboring helices are different for helical pairs with H-bonds and helical pairs without H-bonds. Figure 3 shows the sorted distributions of the number of interhelical atomic contacts. A Wilcoxon rank-sum test strongly suggests that the average numbers of atomic contacts are different for these two populations ($P = 8.1\times10^{-20}$). A Kolmogorov-Smirnov test strongly suggests that the empirical cumulative distribution functions for these two populations are different ($P = 2.9\times10^{-7}$). We can, therefore, conclude that on average helices with interhelical H-bonds are packed tighter and hence have more atomic contacts.

Classes of Interhelical H-Bonds in TM Helices

Among the 299 H-bonds identified from the set of 13 proteins, side chains from residues S, Y, T, and H participate in almost 50% of all identified interhelical H-bonds (Table II). The H-bonds formed between TM helices can be classified into two types (Fig. 4): those between two side chains (type SC) and those between a side chain and a main chain backbone nitrogen or oxygen (type SB). Both types occur with approximately equal frequency (51% of H-bonds are of type SC and 49% are of type SB). Ninety percent of observed SB type H-bonds are between a side chain and a main chain carbonyl oxygen. The frequencies at which side chains are found in these two types of H-bonds are shown in Figure 5. Among these, side chains of R and H residues show stronger preference to form H-bonds of type SB than type SC, while side chains of D, Q, and T residues are more frequently found to participate in type SC H-bonds. There is approximately an equal number of H-bonds of both types involving side chains of S, Y, W, N, and K residues.

The number count of observed H-bonds for each pair of eleven types of amino acid residues that are capable of forming H-bond in side chain contact (type SC) is summarized in Table III. These data show that S residue can form H-bonds of SC type with side chains of all 11 residue types. Residue T forms H-bonds with all possible partners except residue K. Side chains of residues E, H, and Y are observed to form H-bond with side chains of 9 other residues out of 11 (there are no H-bonds of residue E with residues D and W, residue H with residues N and R, residue Y with residues K and self). Residues K, R, and W have the least diversity in interacting partners. In addition, there are eight amino acid residue types whose side chains are capable of forming H-bond with another residue of the same type ("self-pairs"). However, only N, E, H, S, and T residues are observed to form self-paired H-bonds. Side chains of residues S, T, and Y are found in the largest number of H-bonds (43, 35, and 36, respectively), while side chains of residues K, R, and E form the smallest number of H-bonds (9, 20, and 21, respectively). All these observations are based on a small sample size (134 unique helices), and the number of different types of residues is

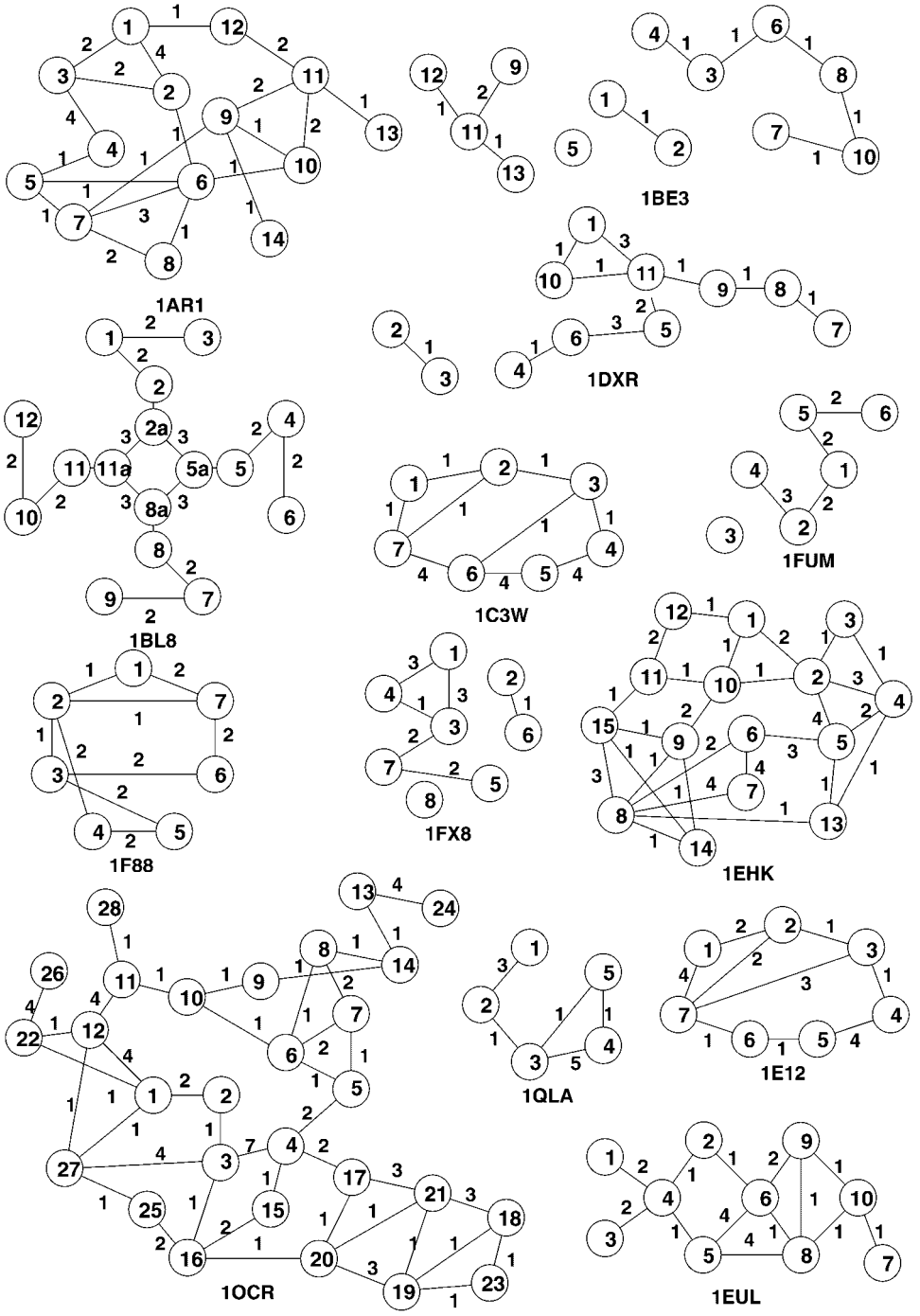

Fig. 2. The topological maps of interhelical H-bonds (top view) of the TM region for each of the 13 proteins. Two helices are connected by an edge if there is at least one H-bond between them. The number of observed interhelical H-bonds is indicated next to the corresponding edge. The helices are numbered by the order in which they appear in the primary sequence.

not equal, but they may be indicative of the nature of H-bonds in membrane protein.

Spatial Motifs of Amino Acids in TM Helices

There are three types of spatial arrangements or clusters of interhelical H-bonds that are observed between two interacting helices. The most frequently observed arrangement has one H-bond and is formed by two amino acid residues from two neighboring helices (H-bond cluster I). They occur at varying positions along the helical pairs. Figure 6(a) shows H-bond between residues W222 and Y305 from subunit I of *Paracoccus dentrificans* cytochrome

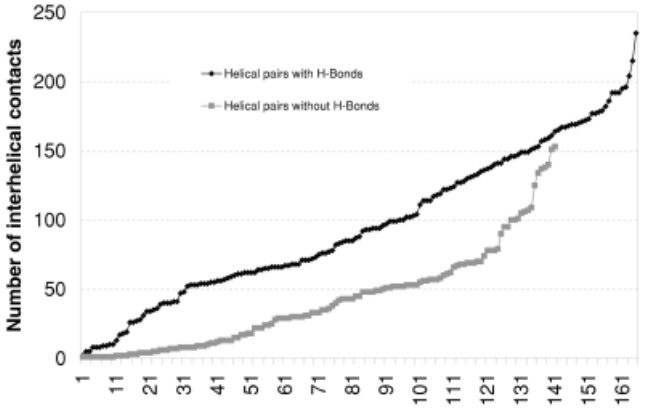

Fig. 3. Interhelical atomic contacts between helical pairs with and without hydrogen bonding. Contacting helical pairs are sorted by the number of atomic contacts on the X-axis, and the atomic contact numbers are plotted on the Y-axis. Helical pairs with H-Bonds in general have much more atomic contacts, and are packed tighter.

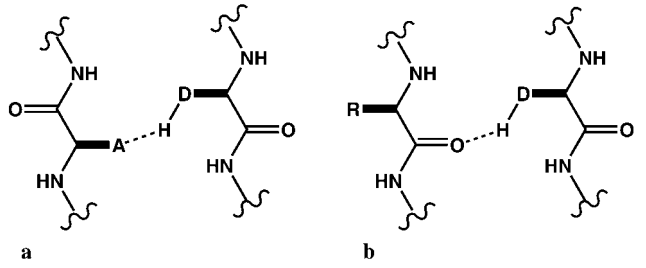

Fig. 4. Two types of H-bonds observed in TM regions of membrane proteins. **a:** Type SC: H-bond is formed between side chains of amino acids from neighboring helices; **b:** Type SB: H-bond is formed between a side chain of an amino acid of one helix and a carbonyl oxygen (depicted) or an amide hydrogen of a polypeptide backbone from a neighboring helix.

TABLE II. Frequencies With Which Amino Acid Side Chains Are Found in H-Bonds

Amino acid	S	Y	T	H	R	W	E	N	D	Q	K	C
%	16	12	10	10	10	9	8	8	7	6	4	1

c oxidase (1ar1), which is located at the ends of the TM helices and near the membrane-solution interface. Figure 6(b) shows H-bond pair between residues S36 and N76 from *Thermus thermophilus* cytochrome c oxidase (1ehk, subunit I), which occurs close to the middles of helices in the hydrophobic core of the phospholipid bilayer. There are helical pairs that contain more than one such cluster, as depicted in Figure 6(c).

"Serine zipper" Motif

The second type, H-bond cluster II, is formed by two amino acid residues with two SB-type hydrogen bonds between them. This is mostly seen for S-S pairs [Fig. 7(a)], e.g., S148-S178 in subunit I of cytochrome c oxidase from *Paracoccus dentrificans*. H-S pair offers another example (cytochrome c oxidase from *Paracoccus dentrificans*; H94-S46). There are two H-bond clusters II on helices III and IV

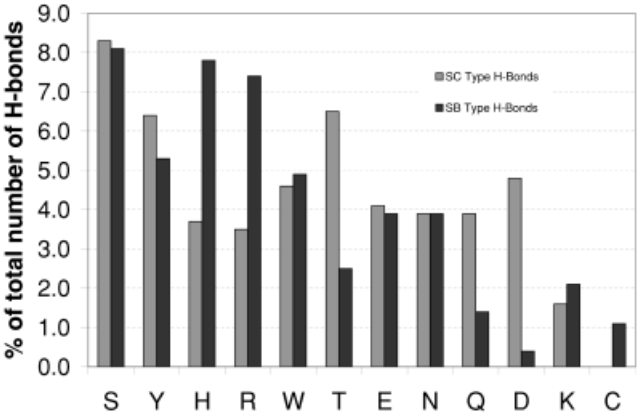

Fig. 5. Percentage of H-bonds of types SC and SB formed by different amino acid residues.

of bovine cytochrome c oxidase (S101-S156 and S108-S149, subunit I) with six intervening amino acid residues between the S residues. In addition, there is a SB type H-bond between residues S115 and S142. The overall spatial arrangement of H-bond clusters II between these two helices is similar to the leucine zipper coiled coil heptad motif.[29] By analogy, these S-S pairs can be thought of as "serine zipper" [Fig. 7(b)], which may facilitate a tight association between helices by providing five or more interhelical H-bonds between helices III and IV of subunit I of bovine cytochrome c oxidase.

To understand how common the "serine zipper" motif appears in cytochrome c oxidases from different organisms, we performed a PSI-BLAST[30] search on non-redundant database of trans-membrane domains.[2] This leads to the identification of 105 non-redundant sequences that have more than 73% sequence identity to helix III. In most of these sequences, S residues occupy positions 7, 14, and 21. Positions 7 and 21 are conserved in all sequences, and Ser at position 14 was conserved in 98 sequences. Analogous search with helix IV leads to 54 sequences from the non-redundant peptide database, where S residues at positions 1, 8, and 15 were conserved in all sequences. Figure 8 shows the sequence logos[31] of homologs of helices III and IV. Sequence logos graphically represent the conservation of residues at various positions in a set of aligned sequences.

Why are these S residues conserved? We found that these residues are tightly packed with other residues in the TM regions. A correlation between the number of interhelical contacts of the residue in the helix and the degree of its conservation among the homologous sequences is found for both helices. The average number of interhelical contacts per residue is 23 for helix III and 18 for helix IV. From the sequence logos, the least conserved residues on helix III are L110, L113, and A114. The number of interhelical contacts for each of these residues is well below the average (4, 2, and 9, respectively). The same observation is true for residues T146 and V155 from helix IV, which are the least conserved, and the number of contacts per residue is well below average (10 and 3, respectively). On the other hand, all serine residues partici-

TABLE III. Observed H-Bonds (Number Count) Between Side Chains of Amino Acid Residues (SC Type)

	ARG	ASN	ASP	GLN	GLU	HIS	LYS	SER	THR	TRP	TYR
ARG	0	0	5	3	5	0	0	4	1	0	2
ASN	0	1	4	0	2	0	2	4	1	3	4
ASP	5	4	0	0	0	1	1	2	3	2	9
GLN	3	0	0	0	1	1	2	8	4	2	1
GLU	5	2	0	1	3	1	2	2	3	0	1
HIS	0	0	1	1	2	1	1	3	6	4	2
LYS	0	2	1	2	2	1	0	1	0	0	0
SER	4	4	2	8	2	3	1	4	4	7	4
THR	1	1	3	4	3	6	0	4	2	3	8
TRP	0	3	2	2	0	4	0	7	3	0	5
TYR	2	4	9	1	1	2	0	4	8	5	0

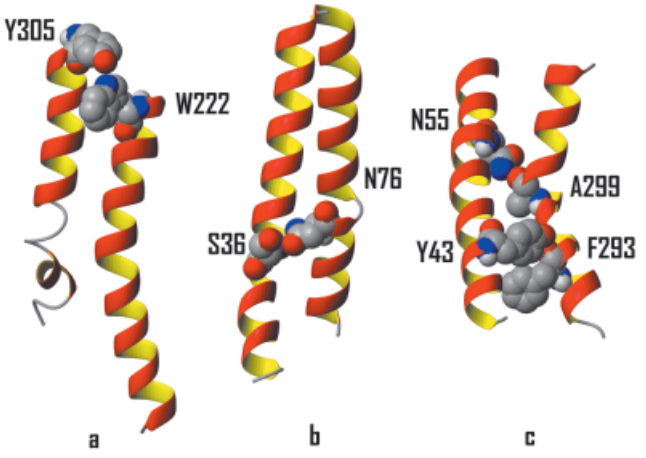

Fig. 6. H-bond cluster I. **a:** Cluster occurs at the ends of TM helices in the phospholipid bilayer interface. Here, hydroxyl oxygen of Y305 accepts NE hydrogen from W222 (subunit I of paracoccus cytochrome c oxidase (1ar1)). **b:** Cluster occurs closer to the middle of the helical pair in the hydrophobic region of phospholipid bilayer. Here, OD1 oxygen of residue N76 accepts hydroxyl hydrogen of residue S36 (subunit I of cytochrome c oxidase from *Thermus thermophilus* (1ehk)); **c:** Two H-bond clusters I in the helical pair between helices I and VII of rhodopsin (1f88). The first cluster is between OH hydrogen of Y43 and carbonyl oxygen of F293 and the second cluster is between ND2 hydrogen of N55 and carbonyl oxygen of A299. All of the molecular structure representations were drawn with the program MOLMOL.[44]

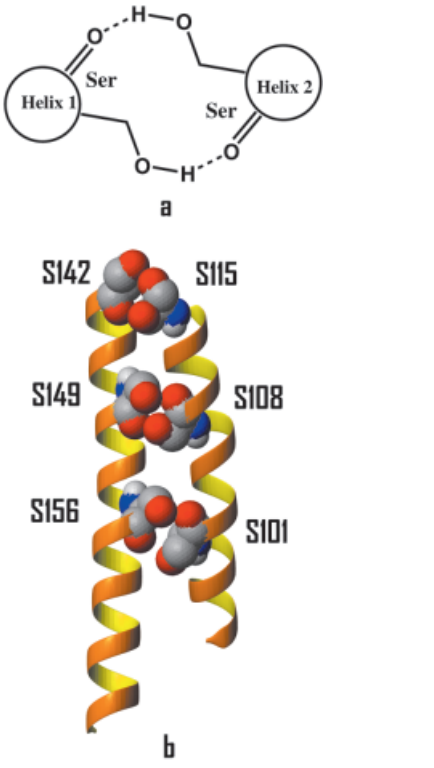

Fig. 7. **a:** Schematic representation of H-bond cluster II between two serine residues on two adjucent helices. Side chain of each S residue forms a H-bond with carbonyl oxygen of its counterpart. There are two SB type H-bonds formed in this cluster; **b:** Serine zipper in bovine cytochrome c oxidase between helices III and IV. H-bond clusters II: S101-S156, S108-S149, and SB type H-bond cluster I: S115-S142. All of the molecular structure representations were drawn with the program MOLMOL.[44]

pating in serine zipper are well conserved, and the numbers of interhelical contacts are near or above the average. For example, residues S101, S108, and S115 on helix III form 20, 32, and 20 contacts, respectively, while on helix IV, residues S142, S149, and S156 form 16, 25, and 18 contacts, respectively.

All these matched sequences belong to cytochrome c oxidases from different organisms. Serine zipper may turn out to be an important structural feature that is evolutionary conserved in this protein family. A recent statistical analysis of amino acid sequence patterns in the TM helices[2] showed that SS7 motif (two S residues separated by 6 other residues) is over-represented in the multi-span SwissProt database containing 12,743 entries with an odds ratio of 1.14 against a random model (P value = 7.48×10^{-4}). We postulate that in many cases this SS7 motif may be a part of the spatial serine zipper, which may occur in other membrane proteins, whose structures are still unknown.

"Polar Clamp" Motif

The third type of the H-bond spatial motif, H-bond cluster III, is formed by three amino acids on two different helices, with two interhelical H-bonds. We call this type a "polar clamp." In most cases, the side chain of an amino acid capable to form at least two hydrogen bonds (i.e., E, K, N, Q, R, S, T) is "clamped" by H-bonds formed with either two side chains or a side chain and a main chain oxygen or

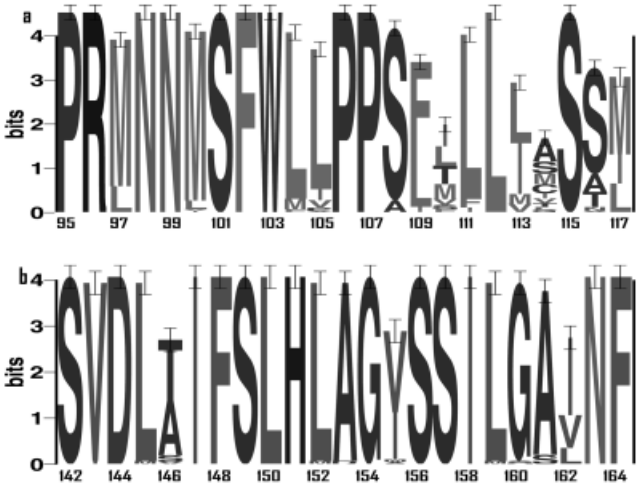

Fig. 8. Sequence logos of homologs of helices III (**a**) and IV (**b**) from bovine cytochrome c oxidase. Numbers on the X-axis are amino acid residue numbers in a protein sequence. Sequence logo graphically represent the conservation of amino acid residues in a set of sequences. These logos were generated from the web site http://www.bio.cam.ac.uk/seqlogo.

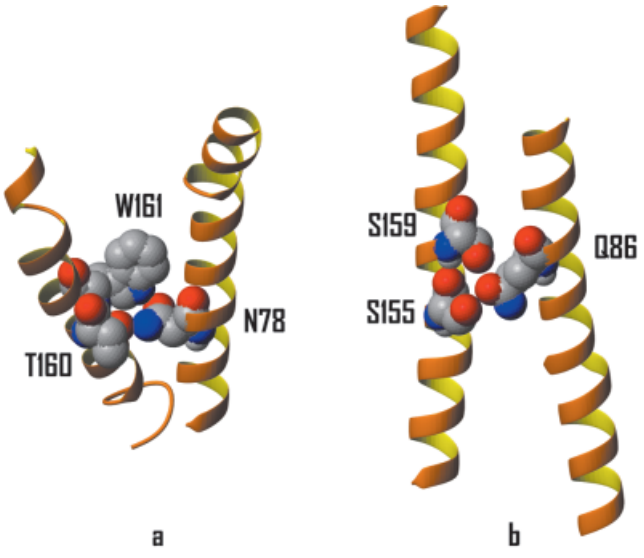

Fig. 9. H-bond cluster III (polar clamp). **a:** A polar clamp in rhodopsin is formed by residues W161 and T160 from helix IV and by N78 from helix II. The side chain of W161 is positioned such that its NE1 atom forms a H-bond with the OD1 atom from N78, while OG1 oxygen from T160 is H-bonded to one of ND hydrogens of N78. **b:** Polar clamp between helices $\alpha 2$ and $\alpha 4$ in subunit I of cytochrome c oxidase from *Thermus thermophilus.* Here, S155 and S159 (positions *i* and *i+4,* helix $\alpha 4$) securely clamp the side chain of Q86 (helix $\alpha 2$). All of the molecular structure representations were drawn with the program MOLMOL.[44]

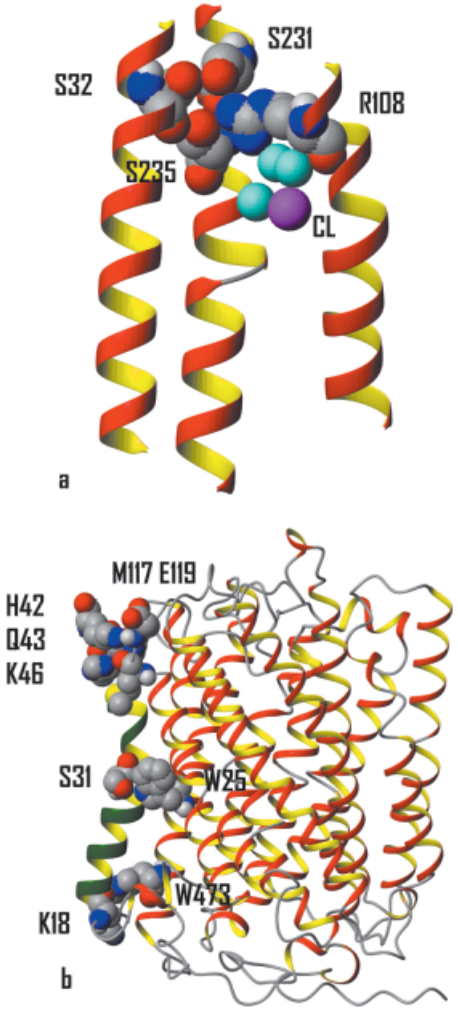

Fig. 10. **a:** Two SS4 polar clamps formed by residues S231 and S235. In the first polar clamp, S32 is clamped by S231 and S235 (Table III): the hydroxyl hydrogens from the side chains of residues S231 and S235 form H-bonds with the carbonyl oxygen and the hydroxyl oxygen of residue S32. In the second polar clamp, hydroxyl oxygens of the same residues form another pair of H-bonds with the guanidinium group of residue R108. This guanidinium group of residue R108 is indirectly connected with chloride ion (shown in magenta) through water molecules (shown in cyan); **b:** The H-bond clusters between subunits I (red and yellow ribbons) and VIIc (green and yellow ribbons) of bovine cytochrome c oxidase. Here, the only TM helix in subunit VIIc forms two H-bond clusters I and two polar clamps with three different helices from subunit I. Two polar clamps are formed by residues H42, Q43, and K46 on helix I (subunit VIIc), which clamp with their side chains the carbonyl oxygens of residues M117 and E119 on helix III (subunit I). Two additional H-bond clusters I are formed between the side chains of residues S31 (helix I, subunit VIIc) and W25 (helix I, subunit I) and between the carbonyl oxygen of residue K18 (helix I, subunit VIIc) and side chain of residue W473 (helix XII, subunit I). All of the molecular structure representations in this figure were drawn with the program MOLMOL.[44]

nitrogen, or two main chain oxygens (nitrogens) of residues at positions *i* and *i+1. . . i+4.* Figure 9 shows two examples of polar clamps. A polar clamp in rhodopsin [Fig. 9(a)] is formed by residues T160 and W161 (positions *i* and *i+1*) from helix IV and by N78 from helix II. The side chain of W161 is positioned such that its NE1 atom forms a H-bond with the OD1 atom from N78, while OG1 oxygen from T160 is H-bonded to one of ND hydrogens of N78. These residues are highly conserved among G protein-coupled receptors (GPCR) and may provide important structural constraints for the helices. Figure 9(b) shows an example of a polar clamp in subunit I of cytochrome c oxidase from *Thermus thermophilus*. It is formed by amino acid residues S155 and S159 at positions *i* and *i+4* on helix $\alpha 4$ and Q86 on helix $\alpha 2$. Here, hydroxyls of S residues securely clamp the side chain carbonyl oxygen of residue Q86 with two SC type H-bonds.

TABLE IV. List of Polar Clamps (H-Bond Clusters III)

Protein	Clamping residues	Clamped residue
1ar1	N486-O, S490-OG	R54-NH1
	I459-O, Q463-NE2	S489-OG
1ocr	N451-OD1, S455-OG	R38-NH2, NE
	H42-NE2, K46-NZ	M117-O
	Q43-NE2, K46-NZ	E119-O
	A74-O, K77-NZ	N170-OD1
	T424-O, Q428-NE2	S454-OG
	Q222-O, L223-O	R156-NH1, NH2
	S142-OG, T146-OG1	S115-O, OG
1e12	S231-OG, S235-OG	S32-O, OG
	Q105-NE2, R108-NH2	S231-OG
	S231-OG, S235-OG	R108-NH2
	D182-OD2, W183-NE1	S133-OG
1c3w	M60-O, L61-O	R7-NH1, NH2
	A126-O, T128-O	R134-NH1, NH2
	S193-OG, E194-OE2	E204-OE1, OE2
1b18	L40-O, S44-OG	S69-OG
	W68-NE1, T72-OG1	Y78-OH
1ehk	S155-OG, S159-OG	Q86-OE1
	F285-O, D287-O	K295-NZ
1eul	S766-OG, S767-OG	N911-ND2
1f88	T160-OG1, W161-NE	N78-OD1
1fx8	H66-NH, L67-NH	E14-OE1, OE2
	A201-NH, M202-NH	E152-OE1, OE2
1fum	V115-O, I118-O	R64-NH1, NH2
1qla	Y188-OH, R189-NH2	D122-O, OD2
	H182-ND1, R189-NE	Q129-O, OE1
1dxr	S262-OG, W266-NE1	D36-OD1

How often do polar clamps occur? We found that polar clamps (total number 28) exist in 12 out of the 13 integral TM proteins in the data set. These are listed in Table IV. The largest number (7) of polar clamps is found in bovine cytochrome c oxidase, which is the largest TM protein in the data set. No polar clamps were found in the structure of bovine cytochrome bc1 complex (1be3), which has a relatively low experimental resolution. Four of the 28 polar clamps are formed by side chains of residue S: Two are formed by S231 and S235 in halorhodopsin (1e12), one is formed by S155 and S159 in cytochrome c oxidase from *Thermus thermophilus* (1ehk), and one is found in calcium transporting ATPase (1eul) (S766 and S767). Two of these polar clamps have three intervening residues, i.e., this motif can be written as SS4 following Senes et al.[2] This pair occurs 1.15 times more often in multi-span Swiss-Prot database of transmembrane domains than it would be expected from random (P value = 1.99×10^{-5})[2]. Polar clamps associated with SS4 motifs may be a common spatial motif that frequently occurs in membrane proteins.

DISCUSSION

The folding of soluble proteins is largely driven by the hydrophobic force.[32] In membrane proteins, once the helices are inserted into the lipid bilayer, it is not clear how important the hydrophobic effect is for helix association. Javadpour et al.[18] suggested that the protein folding problem in membranes can be reduced to the problem of understanding how the TM helices pack. In the TM region, membrane proteins pack differently from soluble proteins.[19] For example, the pattern of interhelical atomic contacts between polar atoms is very diverse in membrane proteins, involving pairs between ionizable residues (salt bridges), ionizable residue and polar residue, as well as polar-polar residues. In contrast, interhelical atomic contacts of polar atoms are exclusively found in residue pairs of two ionizable residues (salt bridges) in control set of soluble proteins. In this study, a detailed examination of polar-polar interactions in the TM helices show that many amino acid residues with polar groups form H-bonds with polar side chains of another residue or with backbone carbonyl atoms. H-bonds are common between TM helices.

What fraction of potential H-bonding groups is satisfied in the TM helices? Because interhelical loops are removed in this study, there may be atoms located at the ends of TM helices whose H-bonding capability is only satisfied when the loops are included. To account for this end effect, we subtract 3.6 residues from each helix so the number of residues capable of forming backbone nitrogen or oxygen H-bond on each helix is reduced by 3.6. We estimate that the fractions of satisfied main chain H-bonds is between 81 and 99% for nitrogen atoms, and 82 to 98% for oxygen atoms among the set of membrane proteins. Similarly, the fraction of satisfied H-bonding groups of the side chains is between 66% (Y residue) and 99% (S residue). There seems to be a negative correlation between the fraction of satisfied H-bonds and the resolution of the protein structure.

Current structural and biochemical data suggests that H-bonds in the TM regions play important roles. For example, extensive mutation studies[33] of calcium transporting ATPase (1eul) showed that the substitution of polar residues, which are not directly involved in Ca^{2+} transport, often resulted in full or partial loss of function. According to the analysis presented in this paper, these residues are capable to form H-bonds. For example, our analysis of the ATPase structure indicates that there is H-bond between K297 and E90, which may be critical for the correct positioning of helix M4. Mutants K297M and K297F resulted in 80–90% loss of maximal Ca^{2+} transport activity, while mutants K297R and K297E lost only 30 and 40% of activity, respectively.[34] Arginine and glutamic acid in both mutants K297R and K297E can substitute lysine and form a H-bond, while neither methionine nor phenylalanine can. Another example is H-bond between residue N101 (OD1) and N796 (ND2) that helps to constrain the position of the side chain of N796, which is a Ca^{2+} binding residue. Mutations of N101 resulted in partial loss of function. The same effect was observed when polar residues S766 and S767 were replaced by residues A or V. These two residues form a polar clamp with the side chain of residue N911. As a result, mutant S766C maintains the capability to form H-bonds, and has a 10–30% increase of Ca^{2+} transport activity in comparison with a wild type protein. These examples demonstrate that in addition to van der Waals interactions, the strong interhelical H-bonds may stabilize helical bundles in polytopic membrane proteins. It is possible that such H-bonds may not be

needed when only a small number of helices is involved and when weak C_a-H—O hydrogen bonds are present, as in the well-studied case of glycophorin A.[11]

H-bonds may also provide specificity for the packing of helices. Analysis of pairwise atomic interhelical interactions in membrane proteins[19] showed that these interactions are dominated by ~20 or so pairs of amino acids with non-polar side chains, suggesting that the background of interhelical interfaces are structurally and chemically rather homogeneous. Dominated by hydrophobic residues (L, I, F, M, and V), there may not be enough specificity without H-bonds for proper packing and positioning of the helices. The structural analysis of membrane proteins revealed three types of spatial clusters of H-bonds in TM helices. In H-bond cluster I, only one H-bond is formed between two amino acid residues. In H-bond cluster II and H-bond cluster III ("polar clamp"), participating residues are arranged in such a way that there are two H-bonds between two and three amino acid residues, respectively.

Polar clamp clusters may be a general spatial motif that is important for membrane proteins. They provide favorable stabilizing interactions, and may enhance the specific orientation of residue side chains necessary for the function of the protein. Detailed examination of polar clamps in halorhodopsin indicate their importance. There are two SS4 polar clamps formed by residues S231 and S235 [see Table IV and Fig. 10(a)]. All residues in these polar clamps (S32, R108, S231, and S235) are well conserved among seven halorhodopsins, as indicated by multiple sequence alignment using CLUSTALW.[35] In the first polar clamp, S32 is clamped by S231 and S235 (Table IV). Hydroxyl hydrogens from the side chains of residues S231 and S235 form H-bonds with the carbonyl oxygen and the hydroxyl oxygen of residue S32 [Fig. 10(a)]. In the second polar clamp, hydroxyl oxygens of the same residues (S231 and S235) form another pair of H-bonds with the guanidinium group of residue R108. This guanidinium group of residue R108 is indirectly connected with a chloride ion through water molecules.[6] This particular polar clamp, therefore, may be functionally important. It was found that the mutant R108K possesses only 2% of activity of the wild type protein and the neutral mutant R108Q is completely defective in chloride transport. The activity of R108Q mutant can be restored by adding guanidinium salts.[36] Here we propose that SS4 polar clamp formed by residues S231 and S235 may constrain the flexible side chain of residue R108 and provide the correct geometry for its guanidinium group. This group, in turn, provides necessary electrostatics forces for the attraction of a chloride ion and forms H-bonds with water molecules (H_2O 22, 24, and 50 in pdbfile 1e12) that are important for the binding of a chloride ion. Mutant R108K also provides electrostatic attraction, but cannot form the necessary H-bond network with water molecules, since it has only one amino group. Upon addition of the guanidinium salts to the neutral mutant R108Q, the guanidinium ion may be trapped by H-bonds formed with side chains of residues S231 and S235 and by the side chain of the carbonyl oxygen of residue Q. This restores the electrostatics of the active site, the correct geometry of the amino groups, and the H-bond network of water molecules and, ultimately, the transport of chloride ions.

The formation of interhelical H-bond may result in tighter association of helices with increased number of interhelical contacts (Fig. 3). However, this may affect the function of a membrane protein adversely. For example, one of the cystic fibrosis phenotypes of cystic fibrosis transmembrane conductance regulator (CFTR) contains an interhelical H-bond, which is formed between mutated residue V232D on TM helix 4 and a wild type residue Q207 on TM helix 3.[37] The structure of this protein is not known and there are two possible consequences of H-bond formation.[37] First, the interhelical H-bond strengthens the preexisting association of TM helices and does not allow any flexibility between them, which may be necessary for the proper functioning of the receptor. Second, it may reorient helix TM3 relative to helix TM4 through the capture of an existing pore-facing residue.[37]

The mechanism of precise assembly of multisubunit, multihelical transmembrane proteins is not well understood. Point mutation V664E in the TM region of neu/erb-2 proto-oncogene, which codes a tyrosine kinase receptor, results in a lethal phenotype due to the constitutive activation of this protein.[38] Polarized FTIR and magic angle spinning NMR spectroscopy revealed that the side chain carboxyl group of residue E664 is protonated and H-bond between two glutamic acid residues promotes formation and stabilization of the dimer and, consequently, the activation of the tyrosine kinase.[39] This raises the question whether H-bonding is normally involved in the assembly and the recognition of subunits in large transmembrane proteins. The analysis of H-bond clusters in bovine cytochrome c oxidase offers interesting information. In this protein, subunit I spans the membrane 12 times, while subunit VIIc spans the membrane only once. The only TM helix in subunit VIIc forms two H-bond clusters I and two polar clamps with three different helices from subunit I. Two polar clamps are formed at the ends of the helices and share the side chain of residue K46: residues H42, Q43, and K46 on helix I (subunit VIIc) clamp with their side chains the carbonyl oxygens of residues M117 and E119 on helix III (subunit I) [Fig. 10(b)]. Here, four H-bonds are formed with only three polar side chains. Two additional H-bond clusters I occur in the middle of the helices and at the opposite end of the helix I (subunit VIIc). The mid-helix H-bond is formed between the side chains of residues S31 (helix I, subunit VIIc) and W25 (helix I, subunit I). Another H-bond is formed between the carbonyl oxygen of residue K18 (helix I, subunit VIIc) and residue W473 (helix XII, subunit I). In summary, the orientation of the single span subunit VIIc relative to subunit I of bovine cytochrome c oxidase is likely to be determined by four H-bond clusters with six H-bonds formed among three different helices from subunit I.

Recent refolding experiments on the bacteriorhodopsin (bR) demonstrated that van der Waals helix-helix packing interactions alone are not sufficient for bR lattice stability.[40] It was also shown that the assembly of bR requires

each of the seven helices, but does not depend on any of the covalent connections provided by the surface loops.[41] These data, together with the above examples of H-bond clusters, imply that interhelical H-bonds are important for stability, assembly, and sometimes functions of polytopic membrane proteins.

NOTE ADDED IN PROOF

Weak C_α-H—O hydrogen bonds are not included in this study. An analysis of their presence and importance in membrane protein can be found in Senes et al.[11]

ACKNOWLEDGMENTS

We thank Dr. C.M. Deber for stimulating conversations. We thank all structural biologists for depositing the coordinates of membrane proteins in the Protein Data Bank.

REFERENCES

1. Arkin IT, Brunger AT. Statistical analysis of predicted transmembrane α-helices. Biochim Biophys Acta 1998;1429:113–128.
2. Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. J Mol Biol 2000;296:921–936.
3. White SH, Wimley WC. Membrane protein folding and stability: Physical principles. Annu Rev Biophys Biomol Struct 1999;28:319–365.
4. Luecke H, Schobert B, Richter H-T, Cartailler J-P, Lanyi JK. Structure of bacteriorhodopsin at 1.55 A resolution. J Mol Biol 1999;291:899–911.
5. Fu D, Libson A, Miercke LJW, et al. Structure of a glycerol-conducting channel and the basis for its selectivity. Science 2000;290:481–486.
6. Kolbe M, Besir J, L.O. E, Oesterhelt D. Structure of light-driven chloride pump halorhodopsin at 1.8 A resolution. Science 2000;288: 1390–1396.
7. Palczewski K, Kumasaka T, Hori T, et al. Crystal structure of rhodopsin: a G protein-coupled receptor. Science 2000;289:739–745.
8. Toyoshima C, Nakasako M, Nomura OH. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. Nature 2000;405:647–655.
9. Choma C, Gratkowski H, Lear JD, DeGrado WF. Asparagine-mediated self-association of a model transmembrane helix. Nature Struct Biol 2000;7:161–166.
10. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM. Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nature Struct Biol 2000;7:154–160.
11. Senes A, Ubarretxena-Belandia I, Engelman DM. The C_a-H. . .O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions. Proc Natl Acad Sci USA 2001;98:9056–9061.
12. Gratkowski J, Lear JD, DeGrado WF. Polar side chains drive the association of model transmembrane peptides. Proc Natl Acad Sci 2001;98:880–885.
13. Shan S, Herschlag D. The change in hydrogen bond strength accompanying charge rearrangement: Implications for enzymatic catalysis. Proc Natl Acad Sci USA 1996;93:14474–14479.
14. Soulimane T, Buse G, Bourenkov GP, Bartunik HD, Hubert R, Than ME. Structure and mechanism of the aberrant ba(3)-cytochrome c oxidase from *Thermus thermophilus*. EMBO J 2000;19:1766–1776.
15. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC. Structure of the E.coli fumarate reductase respiratory complex. Science 1999; 284:1961–1966.
16. Lancaster CRD, Kroeger A, Auer M, Michel J. Structure of fumarate reductase from wolinella succinogenes at 2.2 A resolution. Nature 1999;402:377–385.
17. Fleming KG, Ackerman AL, Engelman DM. The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. J Mol Biol 1997;272:266–275.
18. Javadpour MM, Eilers M, Groesbeek M, Smith SO. Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. Biophysical J 1999;77:1609–1618.
19. Adamian LA, Liang J. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. J Mol Biol 2001;311:891–907.
20. Ostermeier C, Harrenga A, Ermler U, Michel H. Structure at 2.7 A resolution of the Paracoccus denitrificans two-subunit cytochrome c oxidase complexed with an antibody FV fragment. Proc Natl Acad Sci 1997;94:10547–10553.
21. Yoshikawa S, Shinzawa-Itoh K, Nakashima R, et al. Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. Science 1998;280:1712–1713.
22. Edelsbrunner H, Mücke EP. 3-dimensional alpha-shapes. ACM Transactions on graphics 1994;13:43–72.
23. Edelsbrunner H, Shah NR. Incremental topological flipping works for regular triangulations. Algorithmica 1996;15:223–241.
24. Facello MA. Implementation of a randomized algorithm for Delaunay and regular triangulation in 3 dimensions. Comput Aided Geom D 1995;12:349–370.
25. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol 1999;290:253–266.
26. Singh J, Thornton JM. Atlas of protein side-chain interactions, vols. I , II. Oxford: IRL Press; 1992.
27. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
28. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: Structure and implications. Science 1997;276:131–133.
29. Lupas A. Coiled coils: new structures and new functions. Trends Biochem Sci 1996;21:375–382.
30. Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
31. Schneider TD, Stephens RM. Sequence Logos: A new way to display consensus sequences. Nucleic Acids Res 1990;18:6097–6100.
32. Dill K. Dominat forces in protein folding. Biochemistry 1990;29: 7133–7155.
33. MacLennan DH, Rice WJ, Green MN. The mechanism of Ca^{2+} transport by sarco(endo)plasmic reticulum Ca^{2+}-ATPases. J Biol Chem 1997;272:28815–18818.
34. Chen L, Sumbilla C, Lewis D, et al. Short and long range functions of amino acids in the transmembrane region of the sarcoplasmic reticulum ATPase. J Biol Chem 1996;271:10745–10752.
35. Higgins DG, Sharp PM. CLUSTAL: a package for performing nultiple sequence alighment on a microcomputer. Gene 1988;73: 237–244.
36. Rudiger M, Haupts U, Gerwert K, Oesterhelt D. Chemical reconstitution of a chloride pump inactivated by a single point mutation. EMBO J 1995;14:1599–1606.
37. Therien AG, Grant FEM, Deber CM. Interhelical hydrogen bonds in the CFTR membrane domain. Nature Struct Biol 2001;8:597–601.
38. Bargmann CI, Hung M-C, Weinberg RA. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. Cell 1986;45:649–657.
39. Smith SO, Smith CS, Bormann B. Strong hydrogen bonding interactions involving a buried glutamic acid in the transmembrane sequence of the neu/erbB-2 receptor. Nature Struct Biol 1996;3:252–258.
40. Isenbarger TA, Krebs MP. Role of helix-helix interaction in the assembly of the bacteriorhodopsin lattice. Biochemistry 1999;38: 9023–9030.
41. Marti T. Refolding of bacteriorhodopsin from expressed polypeptide fragments. J Biol Chem 1998;273:9312–9322.
42. Iwata S, Lee JW, Okada K, et al. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. Science 1998;281:64–71.
43. Doyle DA, Morais Cabral J, Pfuetzner RA, et al. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 1998;280:69–77.
44. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55.