

Helical Packing Patterns in Membrane and Soluble Proteins

Marina Gimpelev,* Lucy R. Forrest,* Diana Murray,[†] and Barry Honig*

*Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York; and [†]Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York

ABSTRACT This article presents the results of a detailed analysis of helix-helix interactions in membrane and soluble proteins. A data set of interacting pairs of helices in membrane proteins of known structure was constructed and a structure alignment algorithm was used to identify pairs of helices in soluble proteins that superimpose well with pairs of helices in the membrane-protein data set. Most helix pairs in membrane proteins are found to have a significant number of structural homologs in soluble proteins, although in some cases, primarily involving irregular helices, no close homologs exist. An analysis of geometric relationships between interacting helices in the two sets of proteins identifies some differences in the distributions of helix length, interfacial area, packing angle, and distance between the polypeptide backbones. However, a subset of soluble-protein helix pairs that are close structural homologs to membrane-protein helix pairs exhibits distributions that mirror those observed in membrane proteins. The larger average interface size and smaller distance of closest approach seen for helices in membrane proteins appears due in part to a relative enrichment of alanines and glycines, particularly as components of the AxxxA and GxxxG motifs. It is argued that membrane helices are not on average more tightly packed than helices in soluble proteins; they are simply able to approach each other more closely. This enables them to interact over longer distances, which may in turn facilitate their remaining in contact over much of the width of the lipid bilayer. The close structural similarity seen between some pairs of helices in membrane and soluble proteins suggests that packing patterns observed in soluble proteins may be useful in the modeling of membrane proteins. Moreover, there do not appear to be fundamental differences between the magnitude of the forces that drive helix packing in membrane and soluble proteins, suggesting that strategies to make membrane proteins more soluble by mutating surface residues are likely to encounter success, at least in some cases.

INTRODUCTION

The recent increase in the number of membrane proteins whose structures have been solved provides a large data set that can be used in a detailed analysis of the factors that determine their three-dimensional structures (Deisenhofer et al., 1995; Koepke et al., 1996; Tsukihara et al., 1996; Chang et al., 1998; Lancaster et al., 1999; Luecke et al., 1999; Fu et al., 2000; Hunte et al., 2000; Kolbe et al., 2000; Palczewski et al., 2000; Soulimane et al., 2000; Toyoshima et al., 2000; Jordan et al., 2001; Royant et al., 2001; Zhou et al., 2001; Dutzler et al., 2002). One question of considerable interest has been whether there are fundamental differences between the properties of membrane and soluble proteins. For example, as expected, the surfaces of soluble proteins are far more polar than the lipid-interacting regions of membrane proteins. In addition, the fact that the lipid phase does not provide hydrogen bonding partners leads to a strong tendency for membrane proteins to form intramolecular hydrogen bonds, a factor almost certainly responsible for the fact that helices appear as the dominant structural element in this class of proteins. There have been a number of studies that have compared helix packing patterns in membrane and soluble proteins (Bowie, 1997a,b; Eilers et al., 2000, 2002; Adamian and Liang, 2001), and this article adds to that literature.

The large number of structures available for soluble proteins has enabled the construction of increasingly accurate homology models for sequence-related proteins and the development of fold-recognition methods that identify structural homologs even when a sequence signal is weak. Structural genomics initiatives are increasing the database of available structures, and computational methods for structure and function prediction are an essential feature of these large-scale efforts (Burley and Bonanno, 2003; Goldsmith-Fischman and Honig, 2003). Homology modeling has also been widely applied to membrane proteins (Strahs and Weinstein, 1997; Capener et al., 2000; Dwyer, 2001; Becker et al., 2003) and is likely to become more accurate as the database of solved structures increases. One of the goals of the current work is to increase the amount of information available for the structure prediction of membrane proteins by exploiting information available in the large data set of soluble proteins. It is widely recognized that prediction of membrane-protein structure may in many ways be easier than for soluble proteins given the constraints provided by the lipid bilayer and the fact that so many membrane proteins are mostly helical (Chamberlain et al., 2003). Transmembrane (TM) helices are largely, but not exclusively, hydrophobic and consist of stretches of ~15–30 residues (von Heijne, 1994). Since a variety of algorithms are available that predict the location and topology of TM helices with considerable accuracy (Engelman et al., 1986; von Heijne, 1992; Jones et al., 1994; Persson and Argos, 1994; Rost et al., 1995, 1996; Cserzo et al., 1997; Tusnady and

Submitted July 15, 2004, and accepted for publication September 27, 2004.

Address reprint requests to Barry Honig, Dept. of Biochemistry and Molecular Biophysics, Columbia University, 630 W. 168th St., New York, NY 10032. Tel.: 212-305-7970; E-mail: bh6@columbia.edu.

© 2004 by the Biophysical Society

0006-3495/04/12/4075/12 \$2.00

doi: 10.1529/biophysj.104.049288

Simon, 1998), to a first approximation, membrane-protein structure prediction can be viewed in many cases as a problem of packing multiple helices. Knowledge of packing patterns in both membrane and soluble proteins of known structure can provide important information that can be applied to this challenging problem. Significant progress in the ab initio packing of pairs of helices has recently been reported (Fleishman and Ben-Tal, 2002; Kim et al., 2003). The combination of database and biophysical approaches may prove to be particularly effective.

In this article, we report a detailed comparison of helix packing in membrane and soluble proteins. The structure alignment algorithm of PrISM (Yang and Honig, 1999) was used to search the Protein Data Bank (PDB) (Berman et al., 2000) for pairs of interacting helices in soluble proteins that align well with pairs of helices in membrane proteins. (A database of corresponding groups of helices is available on our website—http://trantor.bioc.columbia.edu/packing_pattern). We analyzed and compared the geometries and packing patterns of four data sets of interacting helix pairs: helix pairs extracted from membrane-protein structures; helix pairs in soluble proteins that were detected by PrISM to be structurally similar to the membrane-protein helix pairs (two sets were constructed; see below) and helix pairs in soluble proteins from a nonredundant subset of the PDB. Our results reveal that there can be striking similarities in the geometric and sequence-based properties of individual groups of helices despite significant differences in the composition of residues in the interfaces of membrane proteins, as compared with soluble proteins. The energetic factors that drive membrane-protein folding, and their relationship to those that drive soluble-protein folding, are discussed and the possibility of using the soluble protein database in the modeling of membrane proteins is considered.

MATERIALS AND METHODS

Data sets

We created four data sets that were used in our subsequent analyses: 1), pairs of interacting helices extracted from 16 transmembrane proteins (see Table 1 and supplementary material); 2), pairs of interacting helices from 610 water-soluble proteins taken from the 25% PDB_SELECT list from April 2002 (Hobohm and Sander, 1994); 3), helix pairs included in the 90% PDB_SELECT list that are structurally similar to the interacting helices in the membrane-protein data set; and 4), a subset of set 3 consisting of the best structurally aligned hit for each TM helix pair from the first group. The PDB_SELECT database contains nonredundant structures from the PDB at different levels of sequence identity. All membrane proteins were removed from the 25% and 90% PDB_SELECT lists. Only structures solved to >3.0 Å resolution were considered in all four data sets and NMR structures were excluded from the analysis. In cases where several structures were available, the highest resolution structure was used. For oligomeric proteins, the best-resolved monomer and all interchain interfaces other than those involved in crystal contacts were included in the survey.

Boundaries for TM helical segments were defined with the DSSP program (Kabsch and Sander, 1983), resulting in a set of 171 TM helices. In

TABLE 1 PDB codes and the description of the proteins analyzed

PDB id	Description	No. of TM helices
1ehk	Aberrant ba_3 -cytochrome <i>c</i> oxidase from <i>Thermus thermophilus</i> (Soulimane et al., 2000)	15
1c3w	Bacteriorhodopsin from <i>Halobacterium salinarum</i> (Luecke et al., 1999)	7
1eul	Calcium ATPase from <i>Oryctolagus cuniculus</i> (Toyoshima et al., 2000)	10
1kpl	ClC Chloride channel from <i>Salmonella typhimurium</i> (Dutzler et al., 2002)	18 (dimer)
1ezv	Cytochrome bc_1 complex from <i>Saccharomyces cerevisiae</i> (Hunte et al., 2000)	12
1occ	Cytochrome <i>c</i> oxidase from bovine heart (Tsukihara et al., 1996)	28
1fx8	Glycerol facilitator from <i>Escherichia coli</i> (Fu et al., 2000)	8
1e12	Halorhodopsin from <i>Halobacterium salinarum</i> (Kolbe et al., 2000)	7
1lgh	Light-harvesting complex II from <i>Rhodospirillum rubrum</i> (Koepeke et al., 1996)	2 (dimer)
1msl	Mechanosensitive ion channel from <i>Mycobacterium tuberculosis</i> (Chang et al., 1998)	2 (pentamer)
1prc	Photosynthetic reaction center from <i>Rhodospseudomonas viridis</i> (Deisenhofer et al., 1995)	11
1jb0	Photosystem I from <i>Synechococcus elongates</i> (Jordan et al., 2001)	32
1j95	Potassium channel from <i>Streptomyces lividans</i> (Zhou et al., 2001)	2 (tetramer)
1qla	Respiratory complex II-like fumarate reductase from <i>Wolinella succinogenes</i> (Lancaster et al., 1999)	5 (dimer)
1f88	Rhodopsin from <i>Bos taurus</i> (Palczewski et al., 2000)	7
1h68	Sensory rhodopsin II from <i>Natronomonas pharaonis</i> (Royant et al., 2001)	7

four cases irregular TM helical regions that were broken up into separate segments by DSSP were treated as individual helices since they appear to span the membrane as a continuous secondary-structure element. Each of the membrane-protein structures was divided into interacting helix pairs. Two helices were considered to be interacting if three or more residues of each helix were in contact (Chothia et al., 1981). Two residues were considered to be in contact if the distance between any two of their atoms was within 0.6 Å of the sum of their van der Waals radii, which were derived from the contact distance distributions of 1405 representative protein structures (Li and Nussinov, 1998). Applied to the set of membrane proteins, this procedure yields 265 pairs of helices. The 610 soluble proteins in the 25% PDB_SELECT database contain 3646 helices, which form 2571 interacting helix pairs (termed the “soluble set” in the discussion below).

Angles between helix axes were obtained with the HA2 program (Fleishman and Ben-Tal, 2002), where a helix axis is defined as the line joining the geometrical centers of a set of four $\text{C}\alpha$ atoms at each end of the helix.

Irregular helices

Riek et al. (2001) have pointed out that many TM helices have distinct non- α -helical elements. All TM helices that contain deviations from an α -helical geometry are identified in the supplementary material (Table 1). Four types of irregularities were identified: kink (K); 3_{10} -helix or tight turn (3_{10}); π -helix or wide turn (π); and unwound helix (U). Kinks appear to be

associated with nonideality in neighboring residues that frequently form tight ($K\text{-}3_{10}$) or wide ($K\text{-}\pi$) turns.

The location of kinks in TM and soluble helices were identified using the criterion of Bansal et al. (2000), who noted that when the angle of a local bend in a helix is $>20^\circ$, then the hydrogen bond connecting i and $i + 4$ residues is broken. On this basis, a local bending angle $>20^\circ$ was used to identify kinked helices. The calculations disregarded the four residues at both termini where deviations from ideality often occur. Other deviations from an ideal α -helix were defined using the HELANAL program (Bansal et al., 2000). First, the number of residues per turn and the rise per residue were calculated using a sliding window of four amino acids. Next, each window (or helical turn) was labeled as either α -, 3_{10} -, or π -helix according to the following ranges of values that are based on textbook definitions of helices (Creighton, 1984; Barlow and Thornton, 1988). A turn was defined as α -helical if the number of residues in the turn was in the range between 3.4 and 4.0 and the rise per residue was between 1.36 Å and 1.76 Å. A 3_{10} -helix was defined as having <3.4 residues per turn and a >1.76 Å rise per residue. A π -helix was defined as having >4.0 residues per turn and a <1.36 Å rise per residue. If both criteria were not satisfied the turn was considered to be α -helical.

To eliminate false positives, once the irregular turns had been identified, the dihedral angles of the residues in those turns were calculated using DSSP. Specifically, we defined a turn to be irregular only if at least one of the residues in that turn had dihedral angles in the following ranges (Creighton, 1984; Barlow and Thornton, 1988): $\phi < -66.5$ and $\psi > -29.5$ for a 3_{10} -helix; and $\phi > -59.5$ and $\psi < -55.5$ for a π -helix. Finally, visual inspection of each TM helix was used to verify the reliability of our results and to identify unwound regions of the helices. This process involved looking for obvious deviations from the α -helical backbone structure and hydrogen-bonding pattern. Two 3_{10} -turns were found by inspection to correspond to unwound regions of a helix and one missed π -turn was identified. In all other cases, visual inspection confirmed the identification of the helical irregularities. We were able to identify every irregular helix found by Riek et al. (2001). In addition, we found irregularities in five helices that were not documented by Riek et al., including helices 4, 5, and 7 of the calcium ATPase (1eul), helix 5 of halorhodopsin (1e12), and helix 5 of the photosynthetic reaction center (1prc).

Structural superposition

Structural alignments were carried out with the PrISM program (Yang and Honig, 2000). The groups of interacting membrane-protein helices extracted from the structures in Table 1 ("queries") and listed in the supplementary material (Table 1), were used as substructures to search the 90% PDB_SELECT database with PrISM's structure superposition module. An alignment ("hit") was considered to be significant if at least 75% of the alpha carbon backbone of the query TM helix pair superimposed structurally to within a protein structural distance (PSD) of 0.5 to the soluble helix pair. PSD is a structural similarity measure that accounts for both the relative orientation of secondary structural elements in the two structures and the root mean-square deviation (RMSD), reflecting PrISM's two-stage structure superposition methodology (Yang and Honig, 2000). A PSD of 0.5 is roughly equivalent to an RMSD of 3.5 Å. In general, a number of hits were found for each query, yielding both pairwise and multiple structural alignments. Structure-based sequence alignments were obtained in each case.

Sequence conservation analysis

We analyzed sequence conservation within families of membrane proteins (see below) using ConSurf (Armon et al., 2001; Glaser et al., 2003), a method that employs a physiochemical conservation grade to identify conserved positions in a multiple sequence alignment. The method makes use of phylogenetic trees to ensure that observed levels of conservation are

weighted according to evolutionary distance. PSI-BLAST (Altschul et al., 1997) searches against the nonredundant protein sequence database were used to identify sequence homologs for each membrane-protein structure; the PSI-BLAST hits (after five iterations) along with the seed sequence constitute a "family." An E-value of 1×10^{-5} was used to select hits after all stages of the PSI-BLAST searches (Armon et al., 2001; Glaser et al., 2003). ClustalW (Thompson et al., 1997) was used to create multiple sequence alignments for each family. Multiple alignments were not created for proteins for which PSI-BLAST found five homologs or less. These included: fumarate reductase (1qla), the light-harvesting complex II (1lgh), chain C of the aberrant ba_3 -cytochrome c oxidase (1ehk), chains G and I of cytochrome bc_1 (1ezv), chains I, K, and L of cytochrome c oxidase (1occ), chain H of the photosynthetic reaction center (1prc), and chains I, J, M, and X of photosystem I (1jb0).

Surface area

We calculated the lipid-accessible surface area for the TM residues of all membrane proteins using SURFV (Sridharan et al., 1992). The percentage of residue exposed to the lipid was obtained by dividing the accessible area of the residue in the protein by the area of that residue calculated for the same side-chain conformation within a Gly-X-Gly tripeptide with identical backbone structure. All residues in the TM helices studied here were divided into three groups: interacting residues, defined as residues that have over 80% of their surface area buried in the protein core; partially buried residues, defined as residues that have buried surface areas between 50 and 80%; and residues facing the lipid, defined as residues with more than 50% of their surface area accessible.

RESULTS

Structural alignments

Well-aligned groups of helices

Pairs of interacting helices in membrane proteins that align well with corresponding groups in soluble proteins are listed in the supplementary material (Table 2). An important finding of this study is that 204 out of the 265 interacting TM helix pairs align well with at least one corresponding pair of helices in soluble proteins as defined by a PSD score ≤ 0.5 (RMSDs ranging from 0.5 Å to 3.5 Å). In this way, 5552 pairs of helices from a nonredundant set of soluble proteins were identified and form the third data set to be considered in

TABLE 2 Percentage of helix pairs that contain at least one sequence motif

Data set	Number of pairs	AxxxA (%)	GxxxG (%)	SxxxS (%)
TM	265	38	30	9
TM (<4.5 Å)	60	47	43	7
TM (>4.5 Å)	205	36	26	9
Soluble	2571	32	5	10
Soluble (<4.5 Å)	243	49	14	12
Soluble (>4.5 Å)	2328	30	5	10
Homologous	5552	36	6	8
Homologous (<4.5 Å)	603	51	16	9
Homologous (>4.5 Å)	4949	34	5	8
Best hits	158	40	13	7
Best hits (<4.5 Å)	22	59	9	14
Best hits (>4.5 Å)	136	37	13	6

the analyses below (termed the homologous set). For reference, using a tighter cutoff of 2.0 Å, nearly half (128) of the interacting TM helix pairs have a soluble hit.

The soluble proteins that were identified from the structural alignments belong to 321 different folds within three of the major SCOP classes (Murzin et al., 1995): all α proteins, α/β proteins, and $\alpha + \beta$ proteins. Examples of backbone alignments are shown in Fig. 1. Fig. 1 *a* shows the best structural alignment found between any two TM and soluble helix pairs. The backbones are 100% aligned with a very low RMSD (0.7 Å). Fig. 1 *b* shows the worst alignment obtained within the chosen structural similarity threshold. The higher RMSD (3.5 Å) is clearly reflected by the lower quality of the superposition.

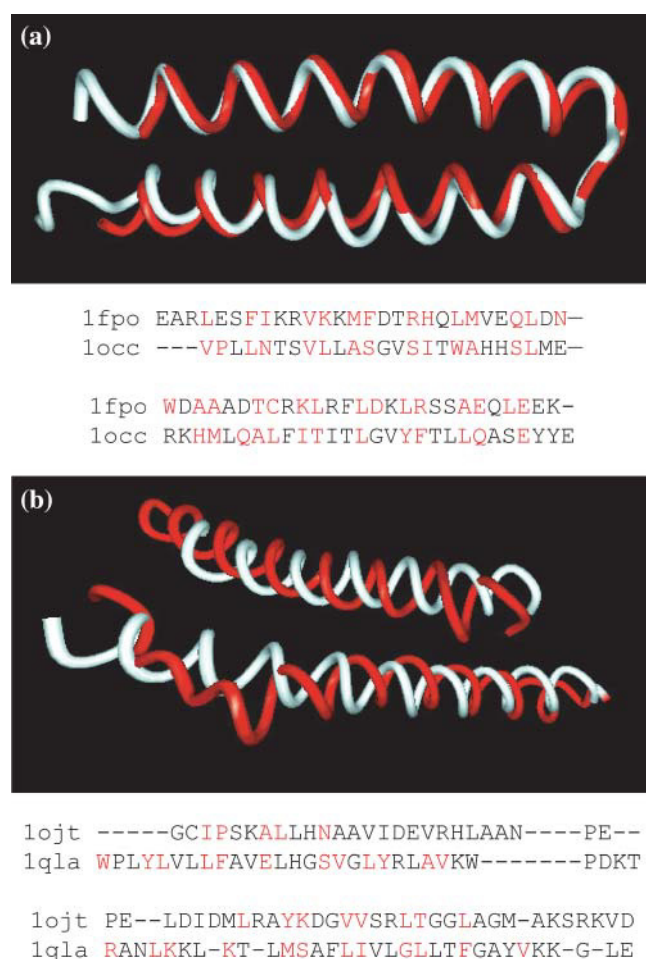


FIGURE 1 Examples of structural alignments between TM and soluble helix pairs. TM helices are shown in white and soluble helices in red. A structure-based sequence alignment of the helices is shown below each superimposition. Contacting residues are highlighted in red. (a) Alignment of cytochrome *c* oxidase (1occ) helices 18 and 19 with chaperone HSC20 molecule (1fpo), residues 110–166. PSD = 0.046, RMSD = 0.7 Å, % alignment = 100%, sequence identity = 7.7%. (b) Alignment of fumarate reductase (1qla) helices 4 and 5 with dihydrolipoamide dehydrogenase (1ojt), residues 165–224. PSD = 0.497, RMSD = 3.5 Å, % alignment = 96.8%, sequence identity = 8.3%.

Irregular helices

Although some TM helix pairs have >100 hits in the database of soluble proteins, 61 helix pairs (30%) have no hits at all. In the latter category, 51 out of the 61 helix pairs have at least one irregular helix and in 26 helix pairs both helices are irregular. In seven of the remaining pairs, the helices are positioned with their N-terminal C α atoms very close together (4–8 Å) and their C-termini far away from each other (>20 Å). Despite the fact that a large fraction of the helix pairs with no hits are irregular, 114 (56%) of the TM helix pairs with hits have at least one irregular helix. In the analysis of the lowest PSD hits for each TM helix pair, in 66% of the cases, kinked helices in TMs are aligned with regular helices in soluble proteins, reflecting the fact that our cutoff is not highly restrictive. In 34% of the cases there are kinked helices in both the membrane protein and its corresponding soluble-protein helix pair.

Examples of well-aligned helices

Fig. 2 illustrates an example of two well-aligned helix pairs (PSD = 0.046; RMSD = 0.7 Å): helices 18 and 19 from cytochrome *c* oxidase (1occ) and helices 7 and 8 from chaperone HSC20 (1fpo). The figure displays the superimposed backbones and the side chains of the interfacial residues. In addition to the backbones, the side chains of the interfacial residues superimpose remarkably well and in some cases have essentially the same conformations. We have found many cases where aligned side chains adopt similar conformations, especially when the residue is the same in both interfaces (see, e.g., Val-121 (1fpo) and Val-136 (1occ) in Fig. 2).

Comparison of helix-helix interfaces

The results of the previous section suggest that soluble proteins contain substructures that have potential predictive value in modeling membrane proteins. In the following sections, we present the results of detailed analyses of helices and helix pairs from membrane proteins (the TM set), from a set of helix pairs taken from the 25% PDB_SELECT list for soluble proteins (the soluble set), from substructures of soluble proteins taken from the 90% PDB_SELECT list of soluble proteins that were found to be structurally similar to TM helix pairs (the homologous set), and from a subset of the homologous set that only includes the best hit to each pair in the TM set (the best-hit set). Some of the comparisons between the TM and soluble sets mirror results reported by Bowie (1997a,b) although given the greater number of structures of membrane proteins now available, our results are based on a larger sample size.

Helix lengths

Fig. 3 shows the distribution of helix lengths for all four sets of helices. There is a strong preference for helices longer

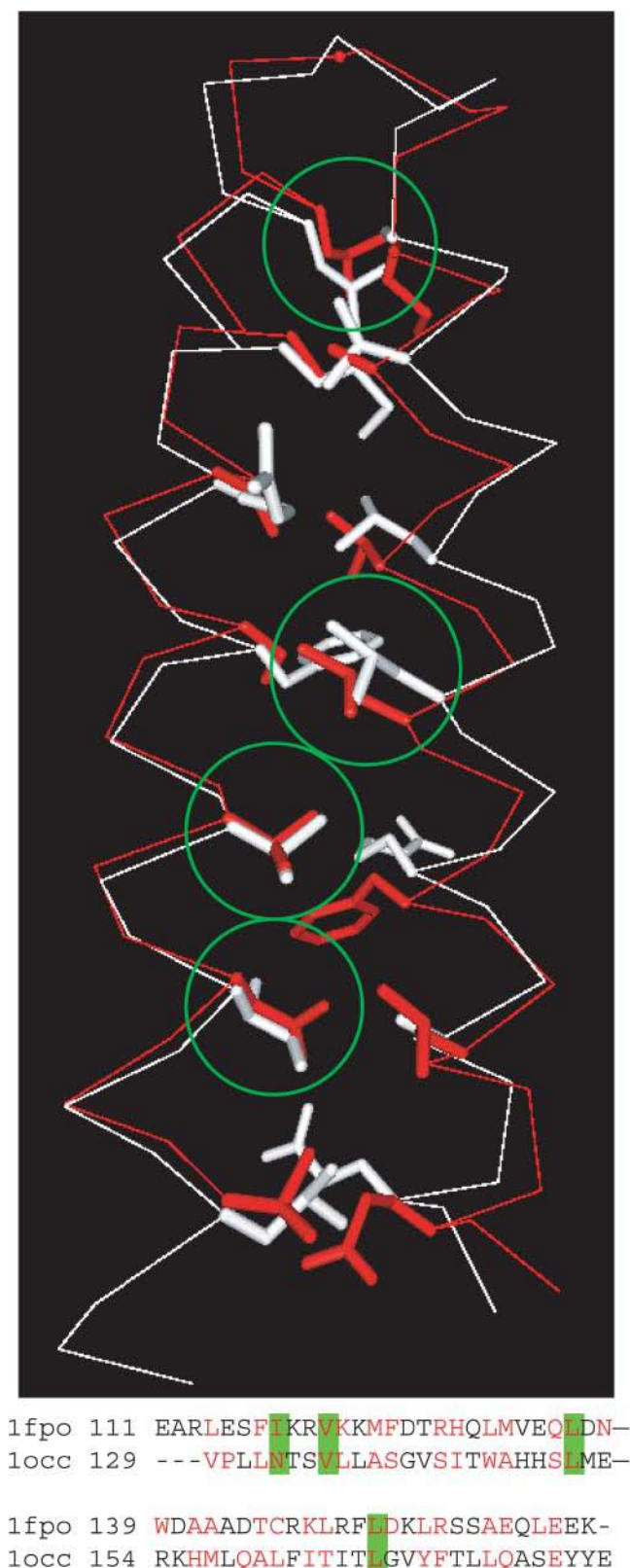


FIGURE 2 Structural superimposition of interacting helices 18 and 19 from the TM protein cytochrome *c* oxidase, 1occ (red) and interacting helices 7 and 8 from the soluble protein chaperone HSC20, 1fpo (white). The side chains of interfacial residues are shown in stick form. A structure-based

than 20 residues in membrane proteins, where we also observe somewhat longer helices than reported by Bowie, with some helices containing as many as 42 residues. Most of the helices in the soluble protein set are between 10 and 19 residues long with an average length of 18 ± 7 residues, compared to an average length of TM helices of 26 ± 6 . Helix lengths in the homologous set are very similar to helices in the soluble set. The distribution of helix lengths for the best-hit set is shifted toward longer helices, as expected, with an average length of 24 ± 9 residues. In contrast to TM helices, there are very few helices in soluble proteins longer than 25 residues, and these belong to the family of coiled coils.

Orientation preferences

We find, in agreement with previous work on membrane proteins (Bowie, 1997b) and water-soluble proteins (Walther et al., 1996), that all sets of proteins favor parallel over antiparallel packing of pairs of helices (see supplementary material). Furthermore, the distribution of helix crossing angles for each of the four groups (Fig. 4) shows a strong preference in TM helices for class *c* packing (Chothia et al., 1981; Walther et al., 1996), which is represented by interhelical angles between 0° and $+30^\circ$. We also find a somewhat narrower distribution of packing angles for TM helices, with no packing angles observed to be $>+69^\circ$ or $<-74^\circ$. In contrast, $\sim 10\%$ of the interacting helices in soluble proteins have packing angles outside of this range. The soluble protein data set and the homologous set show a weaker preference than the TM helices for class *c* packing whereas the best-hit set does show a strong preference for this orientation. Normalized frequencies, as used by Bowie, where the frequencies are divided by the frequencies of the same packing angle for noninteracting helices, were also calculated and found to give the same conclusions (data not shown).

Interactions between adjacent helices

The manner in which two helices that are adjacent in linear sequence space interact with one another can provide useful information in homology modeling. Bowie found that 28 of 32 (88%) pairs of sequential helices interacted structurally, whereas we find 85 of 115 (74%). Thus, with our larger data set, the tendency for sequence-adjacent helices to interact is still quite strong, but weaker than previously observed. For this group of helices, the average loop length between pairs of helices that interact, as well as between pairs of helices that do not interact, is ~ 21 residues. The range of values is such that loop length does not appear to be a good indicator

sequence alignment of the helices is shown below the superimposition, where interfacial residues are highlighted in red. Identical residues that adopt essentially the same conformations are marked in the sequence alignment with green boxes, and on the structure with green circles.

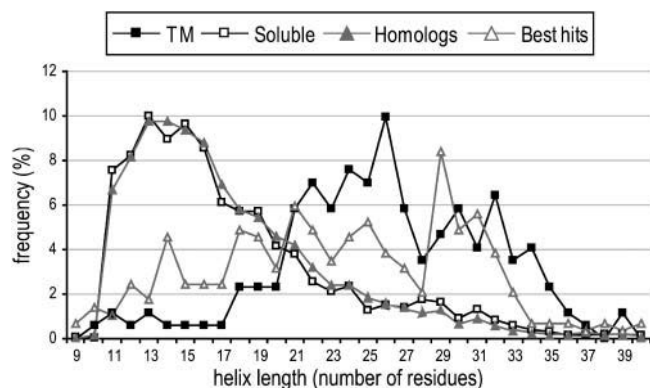


FIGURE 3 Distribution of helix length for individual helices from each of the four data sets: transmembrane helix pairs (TM), the soluble-protein helix pairs (Soluble), soluble-protein helix pairs that are homologous to the TM set (Homologs), and the closest helix pairs from the soluble set to the TM set (Best hits).

of whether two helices will interact; indeed, there are cases where very short loops connect noninteracting helices.

Interhelical distances

As shown in Fig. 5, interacting helices in membrane proteins tend on average to be closer together than interacting helices in soluble proteins. The average of the shortest C α -C α distance between any two helices is 5.5 ± 1.2 Å for membrane proteins, 6.0 ± 1.1 Å for the soluble set, 5.8 ± 1.1 Å for the homologous set, and 5.7 ± 1.1 Å for the best-hit set. As will be discussed below, the fact that interhelical distances are, on average, smaller in TM helices than in soluble proteins does not necessarily imply that membrane proteins are more closely packed (where packing is defined in terms of the volume of cavities between residues) but rather appears to reflect the size of the residues in the helix-helix interface. Indeed, over a third of the interacting helices in soluble proteins have smaller C α -C α distances than the

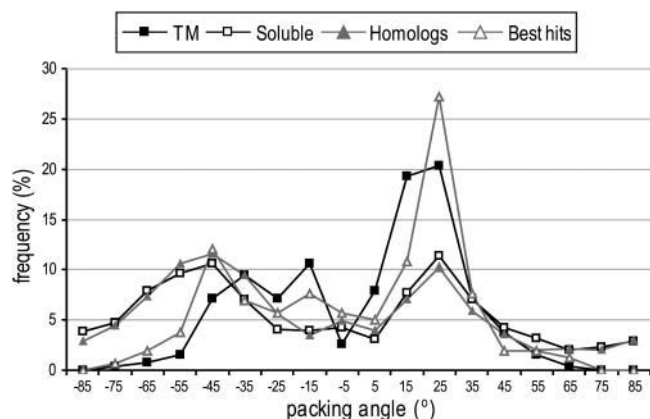


FIGURE 4 Distributions of interhelical packing angles in interacting helix pairs for each of the four data sets (see legend for Fig. 3).

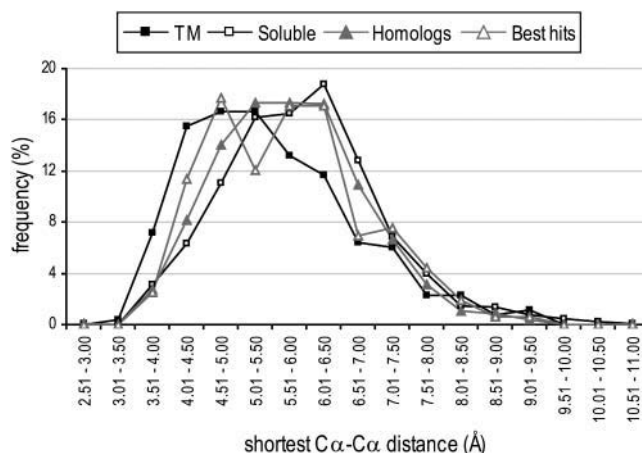


FIGURE 5 Distributions of the shortest C α -C α distances found between interacting helix pairs from the four data sets (see legend for Fig. 3).

average for interacting TM helices. Since the above averages are within one standard deviation, it is feasible that the values for the soluble and TM data sets might approach one another as the number of TM protein structures increases. To check the likelihood of this, an unpaired two-tailed *t*-test was carried out, and the difference between these two data sets was found to be statistically significant, with $p < 0.0001$. As was the case for the packing angles, it is apparent that the PSD cutoff used for the homologous set is not stringent enough to detect a distribution that is significantly shifted from that of the full soluble-protein data set. In contrast, the best-hit set shows a distribution that is similar to that of TM helix pairs.

Irregular helices

The percentage of helices with an irregularity is 40% for the TM set, 19% for the soluble set, 19% for the homologous set, and 32% for the best-hit set. Some helices have more than one irregularity. On a per-residue basis the probabilities for being irregular are 1.9% for the TM set, 1.2% for the soluble set, 1.5% for the homologous set, and 1.8% for the best-hit set. The similarity of these values is surprising and suggests that the well-known propensity of TM helices to be irregular is due, at least in part, to the fact that these helices tend to be longer than those found in soluble proteins. As can be seen in Fig. 6, ~33% of the TM helices, 26% of the best hits, and 11% of each of the soluble and the homologous set helices are kinked. In addition, 1% of the TM helices have an unwound segment. However, as expected, fully unwound helices are not observed in the data set of soluble proteins in this study, since helical regions (including π -, 3_{10} -, and α -helices) were identified using DSSP only.

Interfacial areas

We calculated the surface area buried in the formation of each helix pair by subtracting the surface area of each

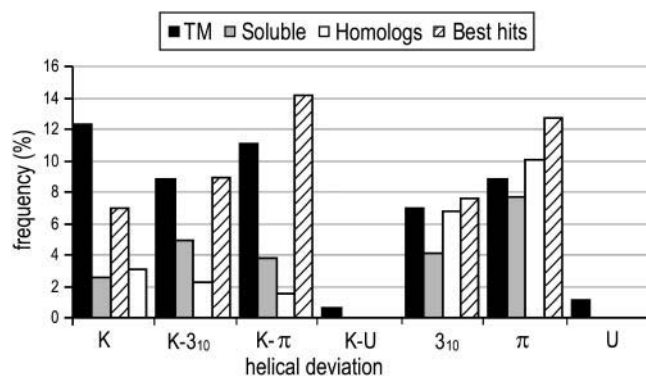


FIGURE 6 The frequency of deviations from α -helicity. Irregularities are as follows: kink (K), tight turn or 3_{10} helix (3_{10}), wide turn or π helix (π), unwound helix (U), kink associated with tight turn ($K-3_{10}$), kink associated with wide turn ($K-\pi$), and kink associated with unwinding of the helix ($K-U$). Data is shown for individual helices from the four data sets of helix pairs (see legend for Fig. 3).

interacting helix pair from the total surface area of the two helices considered alone. Fig. 7 illustrates the distribution of the surface area buried in the helix interfaces of pairs of helices from membrane and soluble proteins. In contrast to that observed for membrane proteins, the distribution of interfacial surface areas for soluble-protein structures is fairly smooth, perhaps due to the larger data set involved. The average buried surface area is $873 \pm 321 \text{ \AA}^2$ for membrane proteins, $667 \pm 285 \text{ \AA}^2$ for the soluble set, $676 \pm 255 \text{ \AA}^2$ for the homologous set, and $818 \pm 363 \text{ \AA}^2$ for the best hits. The fact that TM helix pairs have significantly (to $p < 0.0001$ using an unpaired two-tailed t -test) larger interfacial areas than those observed in soluble proteins is consistent with the fact that TM helices are longer, but it is important to note that this is indeed reflected in larger contact areas.

Residue composition

The results shown in Fig. 8 *a* demonstrate that, for the most part, the distribution of different amino acids in helical inter-

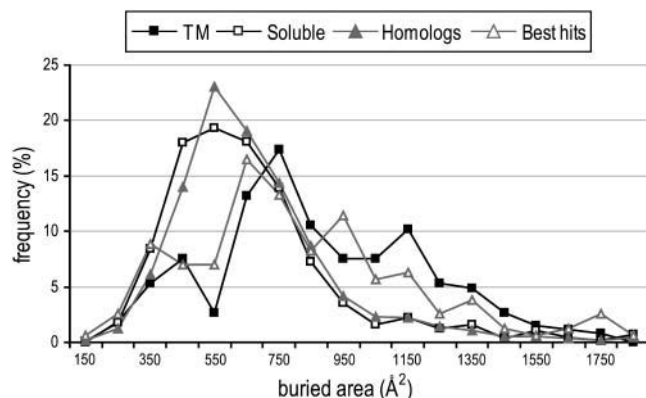


FIGURE 7 Distribution of the surface areas buried in helix interfaces for the four sets of helix pairs (see legend for Fig. 3).

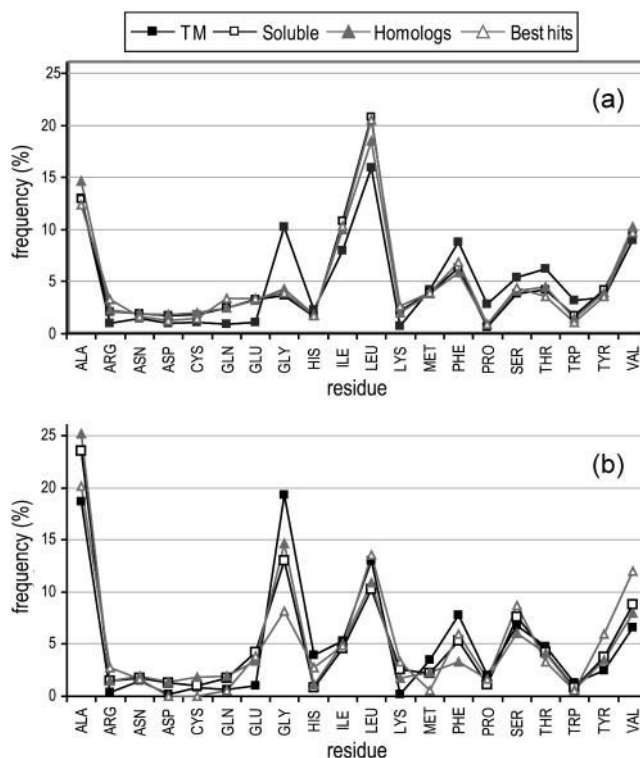


FIGURE 8 (a) The amino acid distributions in interfaces of helix pairs from the four data sets. (b) The amino acid distributions in interfaces of helix pairs with short ($<4.5 \text{ \AA}$) $C\alpha-C\alpha$ distances from each of the four data sets (see legend for Fig. 3). Lines are included for clarity.

faces is quite similar in membrane and soluble proteins. The largest difference between the two distributions is the higher percentage of glycine in membrane proteins, a result that has been found previously (Eilers et al., 2002). Fig. 8 *b* shows the sequence composition of interfaces in pairs of helices with short $C\alpha-C\alpha$ distances ($<4.5 \text{ \AA}$). The distance of 4.5 \AA was chosen as a cutoff because it is ~ 1 standard deviation less than the average distance between helices in TM helix pairs. The number of glycine residues is significantly increased in closely packed helices in all four data sets, and increases are also observed for Ala and Ser residues. It is clear that these three residues, and in particular Gly and Ala, are used to facilitate close packing between helices.

Gly, Ala, and Ser are known to form sequence motifs of the type GxxxG, AxxxA, and SxxxS (Russ and Engelman, 2000; Senes et al., 2000; Kleiger et al., 2002; Liu et al., 2002), which tend to be located in interfacial regions between helices. It is striking that $\sim 1/3$ of the pairs of helices in all four data sets contain at least one AxxxA motif (Table 2). In contrast, the TM set of helix pairs is the only case where a very high proportion contains the GxxxG motif (30%), consistent with the relative enrichment of glycine residues in TM proteins (Fig. 8). It can be seen that the percentage of helix pairs that contain AxxxA and GxxxG motifs increases significantly in all four data sets when considering only those

helix pairs separated by <4.5 Å. The only exception is a small decrease in the number of GxxxG motifs in the best-hit set, perhaps reflecting the small number of pairs of helices, and correspondingly poor statistics, for this data set. In agreement with previous work then, it is apparent that the three sequence motifs facilitate the close packing of α -helices, and that the over-representation of glycine residues leads to increases in the number of closely associated pairs of helices in TM proteins.

However, we also observed cases where closely associated TM pairs with one of the three sequence motifs were able to superimpose almost perfectly on a pair of helices from the soluble set that had no corresponding sequence motifs. In such cases, the side chains of the larger residues located in the interface of helices from the soluble set were oriented sideways, away from the interface, thus allowing very close approach of the two helices. An example of this can be found in the structural alignment of helices 1 and 4 of the glycerol facilitator (1fx8) with helices 7 and 20 of nitroreductase (1f5v), from the best-hit set, which has a C α RMSD of 1.4 Å, and a PSD of 0.085. The Axxx motif in the TM protein aligns structurally with methionine and glutamine residues in the positions corresponding to the alanine residues, but the side chains of the larger amino acids are oriented away from the interface so that the helix backbones are only 4.9 Å apart at their closest point (cf. 4.3 Å in the TM helix pair). More generally, in only 18% of cases does the structural superposition of motif-containing helix pairs from the TM and best-hit set cause the corresponding sequence motifs to be aligned. Thus, although the presence of a motif in a soluble protein facilitates closer approach of the helices, and hence a TM-like packing interaction, the matching of sequence motifs should probably not be considered a reliable modeling heuristic. Nonetheless, the overlap of structural properties between membrane- and soluble-protein helix pairs suggests that soluble proteins may be useful as templates during membrane-protein modeling in that the backbone geometries in the two sets of helices can overlap remarkably well.

Hydrogen bonds

A total of 147 side chain-side chain hydrogen bonds and 133 side chain-backbone hydrogen bonds were identified in the 265 pairs of membrane-protein helices using the geometric criteria of Stickle (Stickle et al., 1992) as implemented in the GRASP2 program (Petrey and Honig, 2003). Briefly, these require an angle of 90 – 180° at the donor atom, and an angle of 90 – 180° or 60 – 180° (for sp^2 and sp^3 hybridized acceptors, respectively) at the acceptor atom; the heavy-atom bond distance must be <3.2 Å, and deviations from planarity of the entire group are allowed for up to 60° or 90° (for sp^2 and sp^3 hybridized acceptors, respectively). Fig. 9 shows the distribution of interhelical hydrogen bonds in helix pairs from membrane and soluble proteins. In both cases there is

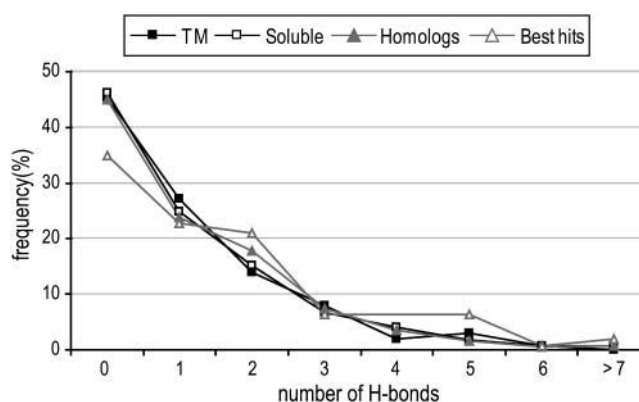


FIGURE 9 Distribution of interhelical hydrogen bonds found in interacting helix pairs from the four data sets (see legend for Fig. 3).

an average of ~ 1 hydrogen bond per pair, although $\sim 1/2$ of the helix pairs have no hydrogen bonds. These numbers are similar to those published earlier for a smaller set of membrane proteins by Adamian and Liang (2002), who noted that every TM helix makes at least one hydrogen bond. Overall, the distribution shown in Fig. 9 is remarkably similar for both membrane and soluble proteins. We find that membrane proteins have $\sim 50\%$ side chain-backbone hydrogen bonds and 50% side chain-side chain hydrogen bonds. Soluble proteins in all three data sets have a higher percentage of side chain-side chain hydrogen bonds (72%, 74%, and 62% for soluble, homologous, and best-hit groups, respectively). The percentage of side chain-backbone hydrogen bonds (70%) is somewhat higher for TM helix pairs whose backbones are closer than 4.5 Å.

Sequence conservation

Fig. 10 summarizes sequence conservation patterns for buried, partially-buried, and lipid-exposed residues among the family members of 14 membrane proteins of known

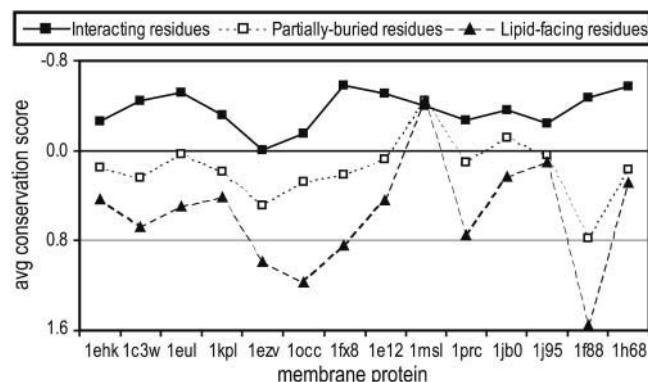


FIGURE 10 Sequence conservation scores averaged over the buried, partially-buried, and lipid-exposed residues in each of the TM proteins, determined using the excessive or surface area (see text for details). Low (negative) scores represent more conserved positions in a protein.

structure (two proteins were excluded because of the small number of known homologs). In all cases, buried residues are the most conserved and partially-buried residues are generally more highly conserved than those that face the lipid bilayer. Many of the buried residues are polar, as might be expected given the important role that polar residues have in driving interhelical interactions in transmembrane helices (Choma et al., 2000; Zhou et al., 2000). Fig. 11 shows the conservation grade for each type of polar amino acid and distinguishes those buried groups that make hydrogen bonds from those that do not.

To identify hydrogen bonds located strictly in the hydrophobic core of the bilayer, we ignored the first and the last turns of every helix. We found 131 hydrogen bonds located in the hydrophobic core in membrane-protein helix pairs, including 67 side chain-side chain and 64 side chain-backbone hydrogen bonds. In 59 out of the 67 side chain-side chain hydrogen bonds, both polar residues are conserved, whereas three have only one conserved residue. Residues with ConSurf scores that are less than zero are considered to be conserved (<http://consurf.tau.ac.il/overview.html>). In 55 out of 64 side chain-backbone hydrogen bonds the residues that supply the side chain are conserved. Thus there is an extremely strong tendency for membrane proteins to conserve buried hydrogen-bonding interactions.

DISCUSSION

A major goal of this study is to provide new data that can help to evaluate similarities and differences between the factors that determine the structures of membrane and soluble proteins. A second goal is to determine whether helix-helix interaction patterns in soluble proteins can be used as templates for the modeling of membrane proteins. Our approach has been to identify helix-helix interaction regions in membrane proteins that are structurally similar to

those found in soluble proteins. Although we have also extended the statistical studies that have been reported previously, our analysis of the subset of globular helix pairs that are structurally similar to membrane-protein helix pairs, as well as our focus on individual cases, offer a novel perspective.

The main conclusion of this work is that most helix-helix interaction patterns seen in membrane proteins also appear in soluble proteins. This suggests that the soluble-protein database might provide a useful resource as templates for modeling helix-helix packing for many TM helix pairs. Most of the exceptions correspond to cases where the TM helix is irregular, a property that appears far more common in membrane proteins than in soluble proteins. As discussed above this appears correlated with the fact that TM helices are longer and therefore have a higher probability of having an irregularity somewhere. Identifying the locations of helical kinks and irregularities poses a serious challenge to homology modeling efforts. The fact that, at least in some cases, similar irregularities can be found in soluble proteins should prove to be helpful in developing prediction procedures. The recent article of Riek et al. (2001) identified some “fuzzy” sequence signals that are characteristic of different types of helical kinks and described changes in axial direction at different types of helical boundaries. It will be of interest to determine whether sequence patterns associated with irregularities in soluble proteins can be used to enhance the statistics for such analysis.

It is clear from our study and from earlier work that similar packing constraints are operable in both soluble and membrane proteins. Grooves-into-ridges rules are largely obeyed and there are no dramatic differences between the packing patterns in the two types of proteins. Those differences that do exist—i.e., TM helices are much longer on average than those in soluble proteins, and there is a narrower distribution of packing angles—appear attributable to the constraints imposed by the lipid bilayer. There are, however, less obvious differences that have been detected previously, some of which are confirmed in the current study. The most notable of these is that the interhelical distances appear to be statistically significantly shorter on average in membrane proteins than in soluble proteins, an effect that appears due in part to the increased presence of small amino acids in the helical interface (Eilers et al., 2002) such as the GxxxG and AxxxA motifs. However, our analysis also indicates that in some cases, close approach of helices in soluble-protein structures is achieved by other means, such as orientation of bulky side chains away from the helix-helix interface.

There is as yet no firm explanation as to why membrane proteins appear designed to have many short interhelical contact distances. One possibility that has been suggested is that tight packing contributes to membrane protein stability. However, it is important to point out that shorter helical contact distances do not necessarily imply stronger van der Waals forces since these are correlated with packing density

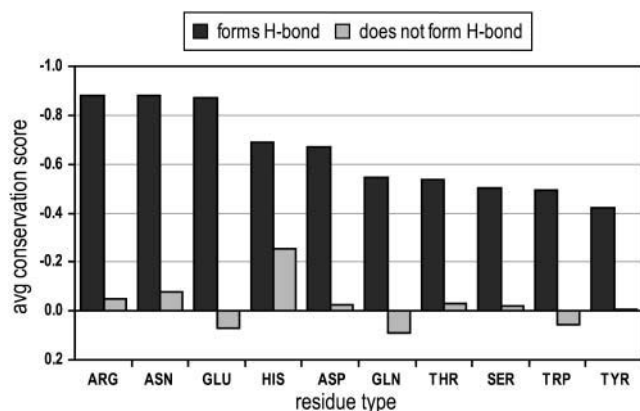


FIGURE 11 Average sequence conservation scores for polar amino acids that form hydrogen bonds and those that do not within the set of TM helices. Low (negative) scores represent more conserved positions in a protein. Residues have been sorted in order of decreasing conservation.

and not the distance between helical axes. Indeed, our observation of a large number of pairs of helices in membrane proteins that are structurally equivalent to those found in soluble proteins suggests that the packing forces are similar in both cases. Bowie and coworkers have recently provided strong evidence that packing interactions provide an important driving force for helices in membrane proteins (Faham et al., 2004). There are apparent differences between the driving forces for interhelical association in membrane and soluble proteins since the hydrophobic effect is only operative in the latter case. This might suggest a stronger driving force for association in the aqueous phase; however, the hydrophobic effect in soluble proteins is largely offset by the free energy penalty associated with desolvating hydrogen-bonded groups as two helices are brought together (Gilson and Honig, 1989). Thus, it may well be that tight packing plays a comparable role in both types of proteins. However, at present this issue remains unresolved. Indeed the role of tight packing in driving the folding of soluble proteins has been difficult to clearly establish and is still an area of considerable uncertainty (see, e.g., Honig, 1999; Liang and Dill, 2001). A similar balance of forces between helices in membrane and soluble proteins is also consistent with the observation that the number of interhelical hydrogen bonds per helix is almost identical for helices in the two environments (Fig. 9, and Adamian and Liang, 2002).

It is interesting to consider the issue of relative packing forces in light of the study of Eilers et al. (2000) who found that helices in membrane proteins have higher packing values than helices in soluble proteins. The data of Eilers et al. (2000) suggests that the results of the two analyses are not mutually inconsistent. A packing value is the fraction of the occluded surface of a residue that is in contact with other residues and the distribution of distances to neighboring atoms (Fleming and Richards, 2000). The distribution of distances is directly related to the packing density but the relationship of buried area to packing density is more complicated. For example, a residue could be 50% buried with its buried area closely packed, whereas another residue might be 75% buried with its buried region less tightly packed against neighboring atoms. Two such residues might have the same packing value but would reflect a different balance of forces. It is interesting in this regard that Eilers et al. find no significant difference in the packing values between residues in membrane and soluble proteins that are >30% buried (Table 4 of Eilers et al., 2000). Rather, the largest difference they detect between water-soluble and membrane proteins is in the fraction of residues that are more than 30% exposed (~16% in membrane proteins and ~25% in soluble proteins). This suggests that the surfaces of soluble proteins are more irregular than those of membrane proteins, but that the interiors of the two classes of proteins are, in fact, very similar in packing density. This, in turn, might reflect the fact that the surfaces of membrane proteins contact alkane chains whereas those of soluble proteins interact with

small water molecules that can more easily penetrate jagged, irregular surfaces.

If stability is the main reason for the close approach of many TM helices, what other causes are suggested? One possibility is simply that close approach in the contact region may allow helices to form larger interfaces, which in turn may facilitate their remaining in contact for a longer distance. This in turn may be required since TM helices must typically span the full width of the lipid bilayer. It was shown above that interacting helices in membrane proteins have larger interfaces on average than in soluble proteins. In Fig. 12 we plot interface size as a function of the shortest C α -C α distance. The strong correlation that is observed suggests that geometric rather than energetic factors may be responsible for the increased number of small residues in helical interfaces in membrane proteins. Finally, it is possible that the smoother helical faces that result from having small interfacial residues may facilitate interhelical motion, which appears important in the function of membrane proteins (see, e.g., Curran and Engelman, 2003).

Our results suggest that helix-helix packing patterns in soluble proteins can be used to sample interhelical conformations that might appear in membrane proteins. Figs. 1 and 2 clearly demonstrate that the backbone geometry of pairs of helices can be nearly identical for membrane and soluble proteins, even if the sequences are quite different. However, in some cases, side-chain conformations can be closely related when the backbones superimpose well (Fig. 2). Thus, in the homology modeling of one membrane-protein sequence onto the structure of another, if the sequences do not align well for a given helical pair it should be possible to use geometrically similar helical pairs taken from the soluble-protein data set to see how different sequences might adapt to a particular orientation. Indeed one could construct structure-based sequence profiles taken from well-aligned helical pairs in the membrane- and soluble-protein data set. At present, the database of membrane-protein structures is too small on its own to allow the use of

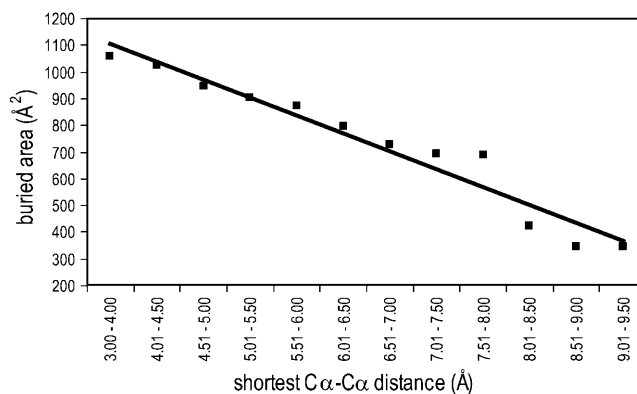


FIGURE 12 Interface size as a function of the minimum C α -C α distance between pairs of helices in TM proteins.

multiple template information as is now possible with soluble proteins (Petrey et al., 2003). The information available in the subset of homologous soluble proteins may also be useful in filtering models of interacting helices that have been generated, for example, using approaches similar to those reported by Bowie and coworkers (Kim et al., 2003). Other filters suggested by this and previous work include sequence-based constraints favoring clusters of conserved residues—in particular those involved in hydrogen bonding—and the presence of sequence motifs in helix interfaces.

Finally, the fact that so many pairs of helices in membrane proteins have close structural homologs in soluble proteins demonstrates that helical interfaces can be quite similar even if the protein surface, and solvent environment, are very different. This in turn suggests that solubilizing membrane proteins by mutating surface residues would not introduce forces that disrupt the internal packing of the protein. Of course small differences in the conformational free energies between the folded and unfolded states have the potential of complicating this strategy in many cases.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

We are grateful to Eric Guoaux, who originally suggested that we use structure alignment to compare helical interactions in membrane and soluble proteins.

Support from the National Science Foundation (MCB-0416708 to B.H. and MCB-0212362 to D.M.) is gratefully acknowledged.

REFERENCES

- Adamian, L., and J. Liang. 2001. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* 311:891–907.
- Adamian, L., and J. Liang. 2002. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins*. 47:209–218.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Armon, A., D. Graur, and N. Ben-Tal. 2001. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.* 307:447–463.
- Bansal, M., S. Kumar, and R. Velavan. 2000. HELANAL: a program to characterize helix geometry in proteins. *J. Biomol. Struct. Dyn.* 17:811–819.
- Barlow, D. J., and J. M. Thornton. 1988. Helix geometry in proteins. *J. Mol. Biol.* 201:601–619.
- Becker, O. M., S. Shacham, Y. Marantz, and S. Noiman. 2003. Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr. Opin. Drug. Discov. Devel.* 6:353–361.
- Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- Bowie, J. U. 1997a. Helix packing angle preferences. *Nat. Struct. Biol.* 4:915–917.
- Bowie, J. U. 1997b. Helix packing in membrane proteins. *J. Mol. Biol.* 272:780–789.
- Burley, S. K., and J. B. Bonanno. 2003. Structural genomics. *Methods Biochem. Anal.* 44:591–612.
- Capener, C. E., I. H. Shrivastava, K. M. Ranatunga, L. R. Forrest, G. R. Smith, and M. S. P. Sansom. 2000. Homology modeling and molecular dynamics simulation studies of an inward rectifier potassium channel. *Biophys. J.* 78:2929–2942.
- Chamberlain, A. K., S. Faham, S. Yohannan, and J. U. Bowie. 2003. Construction of helix-bundle membrane proteins. *Adv. Protein Chem.* 63:19–46.
- Chang, G., R. H. Spencer, A. T. Lee, M. T. Barclay, and D. C. Rees. 1998. Structure of the MscL homolog from *Mycobacterium tuberculosis*: a gated mechanosensitive ion channel. *Science*. 282:2220–2226.
- Choma, C., H. Gratkowski, J. D. Lear, and W. F. DeGrado. 2000. Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.* 7:161–166.
- Chothia, C., M. Levitt, and D. Richardson. 1981. Helix to helix packing in proteins. *J. Mol. Biol.* 145:215–250.
- Creighton, T. E. 1984. Regular conformations of polypeptides. In *Proteins: Structure and Molecular Properties*. W. H. Freeman & Co., New York. 171.
- Cserzo, M., E. Wallin, I. Simon, G. von Heijne, and A. Elofsson. 1997. Prediction of transmembrane α -helices in prokaryotic membrane proteins: the dense alignment surface methods. *Protein Eng.* 10:673–676.
- Curran, A. R., and D. M. Engelman. 2003. Sequence motifs, polar interactions and conformational changes in helical membrane proteins. *Curr. Opin. Struct. Biol.* 13:412–417.
- Deisenhofer, J., O. Epp, I. Sinning, and H. Michel. 1995. Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodospseudomonas viridis*. *J. Mol. Biol.* 246:429–457.
- Dutzler, R., E. B. Campbell, M. Cadene, B. T. Chait, and R. MacKinnon. 2002. X-ray structure of a CIC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature*. 415:287–294.
- Dwyer, D. S. 2001. Model of the 3-D structure of the GLUT3 glucose transporter and molecular dynamics simulation of glucose transport. *Proteins*. 42:531–541.
- Eilers, M., A. B. Patel, W. Liu, and S. O. Smith. 2002. Comparison of helix interactions in membrane and soluble alpha-bundle proteins. *Biophys. J.* 82:2720–2736.
- Eilers, M., S. C. Shekar, T. Shieh, S. O. Smith, and P. J. Fleming. 2000. Internal packing of helical membrane proteins. *Proc. Natl. Acad. Sci. USA*. 97:5796–5801.
- Engelman, D. M., T. A. Steitz, and A. Goldman. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15:321–353.
- Faham, S., D. Yang, E. Bare, S. Yohannan, J. P. Whitelegge, and J. U. Bowie. 2004. Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol.* 335:297–305.
- Fleishman, S., and N. Ben-Tal. 2002. Predicting the conformations of tightly packed alpha-helices in transmembrane proteins. *J. Mol. Biol.* 321:363–378.
- Fleming, P. J., and F. M. Richards. 2000. Protein packing: dependence on protein size, secondary structure and amino acid composition. *J. Mol. Biol.* 299:487–498.
- Fu, D., A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*. 290:481–486.
- Gilson, M., and B. Honig. 1989. Destabilization of an α -helical bundle by helix dipoles. *Proc. Natl. Acad. Sci. USA*. 86:1524–1528.
- Glaser, F., T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, and N. Ben-Tal. 2003. ConSurf: identification of functional regions in

- proteins by surface-mapping of phylogenetic information. *Bioinformatics*. 19:163–164.
- Goldsmith-Fischman, S., and B. Honig. 2003. Structural genomics: Computational methods for structure analysis. *Protein Sci.* 12:1813–1821.
- Hobohm, U. S., and C. Sander. 1994. Enlarged representative set of protein structures. *Protein Sci.* 3:522–524.
- Honig, B. 1999. Protein folding: from the Levinthal paradox to structure prediction. *J. Mol. Biol.* 293:283–293.
- Hunte, C., J. Koepke, C. Lange, T. Robmanith, and H. Michel. 2000. Structure at 2.3 Å resolution of the cytochrome *bc₁* complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure*. 8:669–684.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*. 33:3038–3049.
- Jordan, P., P. Fromme, H. T. Witt, O. Klukas, W. Saenger, and N. Kraub. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*. 411:909–917.
- Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 22:2577–2637.
- Kim, S., A. K. Chamberlain, and J. U. Bowie. 2003. A simple method for modeling transmembrane helix oligomers. *J. Mol. Biol.* 329:831–840.
- Kleiger, G., R. Grothe, P. Mallick, and D. Eisenberg. 2002. GXXXG and AXXXA: common alpha-helical interaction motifs in proteins, particularly in extremophiles. *Biochemistry*. 41:5990–5997.
- Koepke, J., X. Hu, C. Muenke, K. Schulten, and H. Michel. 1996. The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum rubrum*. *Structure*. 4:581–597.
- Kolbe, M., H. Besir, L.-O. Essen, and D. Oesterhelt. 2000. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science*. 288:1390–1396.
- Lancaster, C. R. D., A. Kroger, M. Auer, and H. Michel. 1999. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. *Nature*. 402:377–385.
- Li, A. J., and R. Nussinov. 1998. A set of van der Waals and coulombic radii of protein atoms for molecular and solvent-accessible surface calculation, packing evaluation, and docking. *Proteins*. 32:111–127.
- Liang, J., and K. A. Dill. 2001. Are proteins well-packed? *Biophys. J.* 81:751–766.
- Liu, Y., D. M. Engelman, and M. Gerstein. 2002. Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome Biol.* 3:research0054.
- Luecke, H., B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi. 1999. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* 291:899–911.
- Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of protein database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.
- Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science*. 289:739–745.
- Persson, B. A., and P. Argos. 1994. Prediction of transmembrane segments in proteins utilizing multiple sequence alignments. *J. Mol. Biol.* 237:182–192.
- Petrey, D., and B. Honig. 2003. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequence. *Methods Enzymol.* 374:492–509.
- Petrey, D., Z. Xiang, C. L. Tang, L. Xie, M. Gimpelev, T. Mitros, C. S. Soto, S. Goldsmith-Fischman, A. Kernytsky, A. Schlessinger, I. Y. Y. Koh, E. Alexov, and B. Honig. 2003. Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins*. 53:430–435.
- Riek, R. P., I. Rigoutsos, J. Novotny, and R. M. Graham. 2001. Non- α -helical elements modulate polytopic membrane protein architecture. *J. Mol. Biol.* 306:349–362.
- Rost, B., R. Casadio, P. Fariselli, and C. Sander. 1995. Transmembrane helices predicted at 95% accuracy. *Protein Sci.* 4:521–533.
- Rost, B., P. Fariselli, and R. Casadio. 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* 5:1704–1718.
- Royant, A., P. Nollert, K. Edman, R. Neutze, E. M. Landau, and E. Pebay-Peyroula. 2001. X-ray structure of sensory rhodopsin II at 2.1 Å resolution. *Proc. Natl. Acad. Sci. USA*. 98:10131–10136.
- Russ, W. P., and D. M. Engelman. 2000. The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* 296:911–919.
- Senes, A., M. Gerstein, and D. M. Engelman. 2000. 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.* 296:921–936.
- Soulimane, T., G. Buse, G. Bourenkov, H. D. Bartunik, R. Huber, and M. E. Than. 2000. Structure and mechanism of the aberrant *ba₃*-cytochrome *c* oxidase from *Thermus thermophilus*. *EMBO J.* 19:1766–1776.
- Sridharan, S., A. Nicholls, and B. Honig. 1992. A new vertex algorithm to calculate solvent accessible surface-areas. *FASEB J.* 6:A174.
- Stickle, D. F., L. G. Presta, K. A. Dill, and G. D. Rose. 1992. Hydrogen bonding in globular proteins. *J. Mol. Biol.* 226:1143–1159.
- Strahs, D., and H. Weinstein. 1997. Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors. *Protein Eng.* 10:1019–1038.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.
- Toyoshima, C., M. Nakasako, H. Nomura, and H. Ogawa. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature*. 405:647–655.
- Tsukihara, T., H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, and S. Yoshikawa. 1996. The whole structure of the 13-subunit oxidized cytochrome C oxidase at 2.8 Å. *Science*. 272:1136–1144.
- Tusnady, G. E., and I. Simon. 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* 283:489–506.
- von Heijne, G. 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 225:487–494.
- von Heijne, G. 1994. Membrane proteins: from sequence to structure. *Annu. Rev. Biophys. Biomol. Struct.* 23:167–192.
- Walther, D., F. Eisenhaber, and P. Argos. 1996. Principles of helix-helix packing in proteins: the helical lattice superposition model. *J. Mol. Biol.* 25:536–553.
- Yang, A.-S., and B. Honig. 1999. Sequence to structure alignment in comparative modeling using PrISM. *Proteins. Suppl.* 3:66–72.
- Yang, A.-S., and B. Honig. 2000. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J. Mol. Biol.* 301:665–678.
- Zhou, A. X., M. J. Cocco, W. P. Russ, A. T. Brunger, and D. M. Engelman. 2000. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* 7:154–160.
- Zhou, Y., J. H. Morais-Cabral, A. Kaufman, and R. MacKinnon. 2001. Chemistry of ion coordination and hydration revealed by a K⁺ channel-Fab complex at 2.0 Å resolution. *Nature*. 414:43–48.