Comparison of Helix Interactions in Membrane and Soluble α-Bundle Proteins

Markus Eilers,* Ashish B. Patel,† Wei Liu,* and Steven O. Smith*
Departments of *Biochemistry and Cell Biology and †Physiology and Biophysics, Center for Structural Biology, SUNY Stony Brook, Stony Brook, New York 11794-5115 USA

ABSTRACT Helix-helix interactions are important for the folding, stability, and function of membrane proteins. Here, two independent and complementary methods are used to investigate the nature and distribution of amino acids that mediate helix-helix interactions in membrane and soluble α-bundle proteins. The first method characterizes the packing density of individual amino acids in helical proteins based on the van der Waals surface area occluded by surrounding atoms. We have recently used this method to show that transmembrane helices pack more tightly, on average, than helices in soluble proteins. These studies are extended here to characterize the packing of interfacial and noninterfacial amino acids and the packing of amino acids in the interfaces of helices that have either right- or left-handed crossing angles, and either parallel or antiparallel orientations. We show that the most abundant tightly packed interfacial residues in membrane proteins are Gly, Ala, and Ser, and that helices with left-handed crossing angles are more tightly packed on average than helices with right-handed crossing angles. The second method used to characterize helix-helix interactions involves the use of helix contact plots. We find that helices in membrane proteins exhibit a broader distribution of interhelical contacts than helices in soluble proteins. Both helical membrane and soluble proteins make use of a general motif for helix interactions that relies mainly on four residues (Leu, Ala, Ile, Val) to mediate helix interactions in a fashion characteristic of left-handed helical coiled coils. However, a second motif for mediating helix interactions is revealed by the high occurrence and high average packing values of small and polar residues (Ala, Gly, Ser, Thr) in the helix interfaces of membrane proteins. Finally, we show that there is a strong linear correlation between the occurrence of residues in helix-helix interfaces and their packing values, and discuss these results with respect to membrane protein structure prediction and membrane protein stability.

INTRODUCTION

Membrane and water-soluble proteins commonly fold into bundles of α-helices. However, the nature and distribution of the amino acids in these proteins are very different. The difference in the composition of the surface-exposed residues is well known and simply reflects the environment of the protein, i.e., in soluble proteins polar and charged residues are on the water-accessible surface, whereas in membrane proteins hydrophobic residues cover the lipid-exposed surface (Rees et al., 1989). Much less is known about the nature and distribution of amino acids in the interiors of membrane and soluble proteins.

There is a long history involving efforts to understand the folding and architecture of membrane proteins. The idea that membrane proteins had an "inside-out" architecture (Engelman and Zaccai, 1980) was appealing when it was originally introduced because it provided an explanation for the mechanism of helix association in membrane proteins. The recent analysis of known crystal structures, however, clearly shows that membrane proteins do not have polar cores of amino acids. Rees et al. (1989) showed that the residues in the interior of membrane proteins are less hydrophobic, on average, than the lipid-exposed residues, but are comparable in hydrophobicity to the residues in the interiors of soluble proteins. They proposed the use of a helical "hydrophobic moment" (Eisenberg et al., 1984) as a way to identify the lipid-exposed surface of transmembrane helices. Their analysis left open the question concerning how the hydrophilic residues are distributed in the interior of membrane proteins. More recently, Stevens and Arkin (1999) concluded that the hydrophilic moment is a poor indicator of helix orientation based on an extensive analysis of known membrane protein structures. They found that helical hydrophilic moments did not generally point toward the center of mass of the protein. As a result, there are still unresolved questions involving the internal architecture of membrane proteins.

Over the past few years, we (Javadpour et al., 1999; Eilers et al., 2000) and others (Langosch and Heringa, 1998; Russ and Engelman, 1999; Adamian and Liang, 2001; Ulmschneider and Sansom, 2001) have addressed how helices pack in membrane proteins using a number of different approaches. We have developed two methods for studying helix interactions in membrane proteins. The first method is based on constructing "contact plots" for all interacting helix pairs in a membrane protein (Javadpour et al., 1999). Based on an analysis of four polytopic membrane proteins we showed that glycine had an unusually high occurrence in helix interfaces and at helix crossing points. One limitation of this early study was that few membrane protein structures were available, a situation that has changed considerably over the past three years. Moreover, a detailed comparison

Submitted September 29, 2001, and accepted for publication December 4, 2001.

Address reprint requests to Dr. Steven O. Smith, Center for Structural Biology, Z-5115 138 CMM, Stony Brook, NY 11794-5115. Tel.: 631-632-1210; Fax: 631-632-8575; E-mail: steven.o.smith@sunysb.edu.

0006-3495/02/05/2720/17 $2.00

with soluble proteins was not made. Finally, the focus of this previous study was primarily on the role of glycine in transmembrane helix association. A larger data set allows for a more comprehensive analysis of all amino acids.

The second method we have developed is based on the use of occluded surfaces to probe amino acid packing. We have shown that membrane proteins are generally more tightly packed than helical soluble proteins (Eilers et al., 2000). The packing analysis strongly suggested that small and polar residues contribute to tight helix interactions. However, the origin of the high packing values in membrane proteins could not be unambiguously established because the packing values were not separately calculated for interfacial and noninterfacial residues.

In this paper, we revisit the question of how helix interactions differ between membrane and soluble α-bundle proteins by combining the two methods in our analysis. We restrict our comparison to only those soluble proteins classified as α-bundle proteins, because the α-bundle architecture is most similar to that of membrane proteins and consequently provides the best comparison. With the recent structure determination of several large membrane proteins and the inclusion of α-bundle domains in soluble proteins, the current data set is significantly larger and the average resolution of structures is higher than that previously used for analyzing helix packing (Eilers et al., 2000). We have generated helix contact plots for 11 unique helical membrane proteins and 23 soluble α-bundle proteins and α-bundle domains. As a result, the contact plots now allow us to address helix packing as a function of the location of any residue. In addition, the helix pairs that are defined in our analysis can be categorized as having either left- or right-handed crossing angles and parallel or anti-parallel orientations. This allows us to address differences in packing and helix interactions as a function of the helix geometry.

By combining the packing and helix contact analyses, we are able to show that small and polar residues serve to mediate tight helix-helix interactions in membrane proteins, and propose that these residues constitute a general packing motif that is well-represented in helical membrane proteins. The results refine how the hydrophilic moment of transmembrane helices relates to the internal architecture of membrane proteins, namely that the hydrophilic moment points between helix pairs rather than toward the center of mass of the protein. Finally, we discuss the use of packing values and interfacial propensities for predicting the relative orientation of transmembrane helices.

METHODS

Helix packing: method of occluded surfaces

The occluded surface (OS) method for analyzing packing interactions in proteins has previously been described (Pattabiraman et al., 1995; DeDecker et al., 1996). The OS method calculates packing values at the level of individual atoms, amino acids, or entire proteins. Packing values range from 0.0 to 1, corresponding to totally exposed and totally occluded environments. Hexagonally packed spheres have a maximum packing value of 0.8 due to the void space that exists where the spheres are not in direct contact (Richards and Lim, 1993). The concept behind the OS method is illustrated in Fig. 1. In the OS calculation, van der Waals surfaces are drawn around each atom in the protein and normals are constructed that extend outward until they reach another surface or a length of 2.8 Å, the diameter of a water molecule. The cutoff of 2.8 Å between amino acid surfaces accounts for the possibility that water can occupy that space and therefore the corresponding surface is defined as being nonoccluded (by another amino acid, chromophore, or prosthetic group). The definition of the OS packing value (Fleming and Richards, 2000) takes into account the normalized occluded (or buried) surface area weighted by the distance to the occluding neighbors. The OS packing value (PV) for each residue is defined as

$$PV = \frac{\sum_{\text{atom}}^{\text{res}} [S_O * \langle 1 - RL \rangle_{\text{atom}}]}{S_t}$$

where S_O is the occluded surface area, S_t is the total surface area (sum of occluded and nonoccluded areas), and RL (ray length) is the length of the extended normal divided by 2.8 Å.

The packing values for individual residues can be directly compared because division by the total molecular surface area normalizes the packing value to account for the various sizes of the amino acids. Moreover, the method works equally well for both buried residues and surface residues. We have seen no systematic bias in packing values based on residue size or interfacial or noninterfacial location. This is illustrated in Fig. 2, which plots the difference in the average packing values for amino acids in interfacial or noninterfacial positions of membrane proteins. The packing value differences are all positive, indicating that the interior positions are more tightly packed. Also, there is no significant difference between the packing value differences for the abundant small residues (e.g., Gly, Ala, Ser, Thr) and large residues (e.g., Phe, Trp, Tyr). (The packing value differences are more variable for charged and highly polar residues that are not abundant in the transmembrane helices of membrane proteins.) This is important in the analysis because amino acids with small volumes tend to have high packing values in transmembrane helices, and a question that immediately arises is whether this results from helix-helix interactions or from small residues being surrounded by large residues on the same helix.

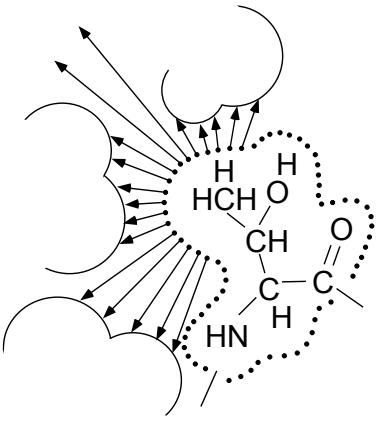

FIGURE 1 Schematic diagram illustrating the occluded surface calculation for the methyl group of threonine. Normals are drawn from the van der Waals surface of the methyl group, and are considered occluded if they encounter another surface within 2.8 Å.

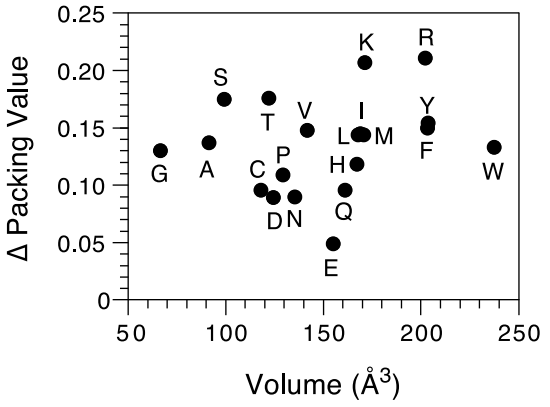

FIGURE 2 Packing differences between interfacial and noninterfacial residues in transmembrane helices as a function of residue volume. The amino acid volumes were taken from Chothia (1975). The interfacial and noninterfacial amino acids in the 11 membrane proteins studied were determined on the basis of helix contact plots.

The occluded surface calculations were carried out on full protein structures, but the packing values we report represent only the amino acids in helices. The average occluded surface packing value for a protein is the average of all of the individual amino acid packing values for that protein. Prosthetic groups and chromophores were included in the calculations, whereas detergent, lipid, and water molecules were excluded. The calculations were carried out on monomers except for the ion channels and the light harvesting complex, where the functional tetramer (1J95), pentamer (1MSL), and trimer (1KZU) were used. We describe in the database section below how the helices in membrane and soluble proteins were assigned.

Helix contact plots

Helix-helix contacts were evaluated using a modified version of the program Euler, which calculates backbone-to-backbone distances between transmembrane helices (Javadpour et al., 1999). The program calculates the interatomic distances between all backbone atoms of each interacting helix pair. In our analysis, two helices were considered to be interacting if the minimum backbone-to-backbone distance was between 3 Å and 8 Å and there were at least 100 distances of <8 Å between backbone heteroatoms. The helix-helix interface is defined by those residues that occur at a local minimum in the contact plots or within 0.5 Å of a local minimum. Noninterfacial residues are those that do not satisfy these criteria and include those that are oriented toward lipids, internal aqueous pockets, or channels in membrane proteins and toward water in soluble proteins. Since our definition of interacting helices covers a broad range of interhelical distances, we separately characterize the interfacial amino acids that have a backbone-to-backbone separation of ≤ 6 Å, and those that have a backbone-to-backbone separation of >6 Å.

Fig. 3 illustrates the concept behind the construction of helix contact plots for three different helix orientations. Fig. 3 *A* presents the contact plot for helix 4 in subunit L interacting with helix 4 in subunit M of the bacterial photosynthetic reaction center (1AIJ). Both subunits have five transmembrane helices, and three of these helices make contact with the other subunit. Helix 4 in both subunits is in the central position of the intersubunit contacts. The two helices cross in the middle of the membrane, and there are three amino acids (Phe-180, Asn-183, and Ala-184), which lie in the ≤ 6 Å interface. Fig. 3 *B* presents the contact plot for helices 2 and 7 in bacteriorhodopsin (1C3W). The two helices diverge at the level of Pro-50 on helix 2. The retinal chromophore of bacteriorhodopsin is attached to Lys-216 on helix 7, which is adjacent to Pro-50. As a result, the open region between helices 2 and 7 may be of functional importance in forming

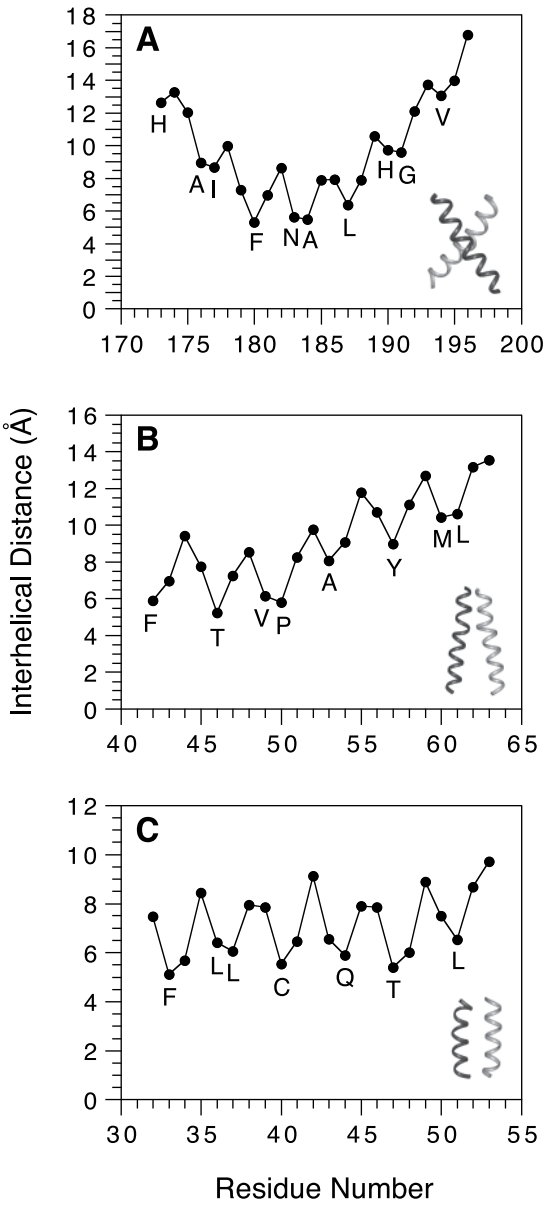

FIGURE 3 Schematic diagram illustrating the helix contact plot analysis. (*A*) Contact plot for helix 4 in subunit L and helix 4 in subunit M of the bacterial photosynthetic reaction center (1AIJ). (*B*) Contact plot for helices 2 and 7 in bacteriorhodopsin (1C3W). (*C*) Contact plot for helices 1 and 2 in cytochrome *b* of the cytochrome *bc1* complex (1BE3).

the retinal binding site. Fig. 3 *C* presents the contact plot for helices 1 (αA) and 2 (αB) in cytochrome *b* of the cytochrome *bc1* complex (1BE3). These two helices coil in a left-handed geometry, form close contacts along their entire length, and serve as part of the scaffold for coordinating hemes b_L and b_H in cytochrome *b*. The helix crossing angles used to characterize left- and right-handed helix pairs for the analysis in Table 3 were calculated using the program define_structure (Richards and Kundrot, 1988).

Statistics: test of significance

The *z*-test was used to evaluate whether the calculated differences in the average (mean) packing values are significant or simply result from sample

TABLE 1 Amino acid packing values in helical membrane and soluble α-bundle proteins

	Membrane Proteins		Soluble α-Bundle Proteins		
	Occurrence (%)	Packing Value	Occurrence (%)	Packing Value	*z*-Test
Ala	10.812	0.488	11.829	0.472	—*
Arg	1.936	0.392	6.375	0.351	+†
Asn	1.841	0.463	3.456	0.371	++‡
Asp	1.180	0.432	4.301	0.364	+
Cys	1.275	0.475	1.229	0.497	—*
Gln	1.180	0.450	5.184	0.346	++
Glu	1.794	0.425	7.028	0.320	++
Gly	7.602	0.524	3.533	0.466	++
His	2.502	0.473	2.650	0.399	++
Ile	8.876	0.412	5.837	0.482	++
Leu	15.014	0.404	14.209	0.465	++
Lys	1.605	0.383	5.645	0.299	++
Met	4.721	0.447	2.957	0.428	—*
Phe	9.537	0.408	3.111	0.491	++
Pro	2.314	0.507	2.035	0.349	++
Ser	5.052	0.474	4.186	0.421	++
Thr	6.185	0.454	4.301	0.411	++
Trp	3.447	0.419	1.651	0.476	++
Tyr	3.494	0.413	3.725	0.455	+
Val	9.632	0.424	7.258	0.470	++
Average		0.441		0.418	++
Standard deviation		0.116		0.128	

*−, $p \geq 0.05$.
†+, $0.01 \leq p < 0.05$.
‡++, $p < 0.01$.

variability. The *z*-test evaluates the difference in the mean values of two sets of data based on the number of elements in the data set and the standard deviation between the elements. We applied the null hypothesis to compare the average packing values. The null hypothesis gives probabilities that the difference of mean values between two populations originates from sample variability. If the difference is significant and does not result simply from large sample variability, then the *p*-values are low. *P*-Values of <0.05 indicate that there is a >95% probability that the difference is significant. In Tables 1 and 2 we divided the results of the *z*-test into three classes: $p \geq 0.05$ (not significant), $0.01 \leq p < 0.05$, and $p < 0.01$, marked as —, +, and ++, respectively.

Database of membrane and soluble proteins

The database used for our analysis included 11 membrane and 23 soluble α-bundle proteins. The 11 transmembrane proteins are all α-helical and of known structure with a resolution of ≤3.5 Å. Only nonhomologous proteins were used in the analysis. When several crystal structures of homologous proteins were present in the Protein Data Bank, the highest-resolution structure was chosen. We included the cytochrome *bc1* complex from *Saccharomyces cerevisiae* (1EZV) rather than either of two lower-resolution structures (1BE3 and 1BCC). Similarly, the photosynthetic reaction center from *Rhodopseudomonas viridis* (1DXR) was included rather than the reaction center from *R. sphaeroides* (1AIJ) or a lower resolution structure from *R. viridis* (6PRC). In the case of the two structures of the light-harvesting complex (1KZU and 1LGH) whose structures have similar resolution (2.5 Å and 2.4 Å, respectively) and R factors (22% and 21%, respectively), we selected the complex from *R. acidophila* (1KZU) whose transmembrane helices have much lower thermal factors (16.1 vs. 31.5, respectively). We included bacteriorhodopsin (1C3W) and excluded halorhodopsin (1E12) because these proteins have the same architecture and their functions can be interconverted by a single amino acid substitution. We did not include membrane-associated proteins (e.g., Lpp-56) or membrane proteins that did not include transmembrane helices (e.g., TolC). The data set was selected from those structures deposited in the Protein Data Bank as of July 2001.

The Appendix lists the resolution and packing values of the proteins used in our analysis. Given the limited data set for helical membrane proteins, we have chosen to include membrane proteins whose resolution ranges from 1.55 Å (1C3W) to 3.5 Å (1MSL). The wide variation of resolution has the potential to influence the conclusions drawn about membrane protein packing and interhelical contacts. However, the higher-resolution structures are generally associated with higher packing values. The four membrane protein structures with a resolution of 2.2 Å or less have packing values well above the average packing value for soluble proteins. Moreover, the resolution of even the 3.5 Å structure of the mechanosensory channel (1MSL) is sufficient to define the relative orientation of the transmembrane helices for the helix-packing analysis. As indicated above, to account for some uncertainty in atomic positions, we define interfacial residues as those that occur at a local minimum in the contact plots or within 0.5 Å of a local minimum.

For comparison with membrane proteins, we analyzed the family of soluble proteins classified as α-bundle proteins. These helical soluble proteins are the most similar in architecture to membrane proteins, and consequently provide the best database for comparison. The soluble proteins selected have known x-ray structures and show no homology. We selected proteins classified as α-bundle proteins in the CATH database as of May 2001 (http://www.biochem.ucl.ac.uk /bsm/cath_new/index.html; Michie et al., 1996; Orengo et al., 1997), and used only those proteins that have at least three helices of nine or more residues. The CATH classification of α-bundle proteins are those that "must have at least 60% α and less than 5% β secondary structure assignment, with at least 50% α-α and less than 5% β-β secondary structure contacts. The helices lie approximately parallel or antiparallel to one another. Specifically, pairwise angles between the helical axes of approximately 0° and 180° predominate."

TABLE 2 Amino acid packing values in membrane and soluble α-bundle proteins

	Membrane Proteins				Soluble α-Bundle Proteins					
	Interface ≤ 6 Å			Noninterface	Interface ≤ 6 Å			Noninterface	*z*-Test Packing Value	
	Occurrence (%)	Propensity	Packing Value	Packing Value	Occurrence (%)	Propensity	Packing Value	Packing Value	Interface ≤ 6 Å	Noninterface
Ala	15.192	1.405	0.539	0.402	19.55	1.726	0.537	0.368	—*	+
Arg	1.327	0.686	0.499	0.288	4.39	0.688	0.399	0.300	0†	—
Asn	2.212	1.201	0.494	0.404	2.53	0.731	0.476	0.318	—	0
Asp	1.032	0.875	0.498	0.410	1.73	0.402	0.405	0.327	0	0
Cys	1.770	1.388	0.491	0.396	2.13	1.731	0.513	0.378	—	0
Gln	0.737	0.625	0.493	0.397	3.06	0.590	0.442	0.309	0	0
Glu	1.327	0.740	0.461	0.412	1.99	0.284	0.410	0.287	0	0
Gly	12.094	1.591	0.570	0.440	3.46	0.979	0.522	0.409	+‡	—
His	1.770	0.707	0.581	0.463	3.32	1.255	0.452	0.312	++§	0
Ile	6.195	0.698	0.478	0.333	6.78	1.162	0.497	0.430	—	++
Leu	11.504	0.766	0.459	0.315	15.82	1.114	0.505	0.367	++	++
Lys	1.180	0.735	0.456	0.249	2.66	0.471	0.340	0.276	0	—
Met	4.720	1.000	0.490	0.346	2.39	0.809	0.497	0.338	—	—
Phe	7.080	0.742	0.473	0.323	2.79	0.898	0.520	0.381	+	—
Pro	4.130	1.785	0.526	0.418	1.99	0.980	0.437	0.298	+	0
Ser	7.375	1.460	0.534	0.359	3.99	0.953	0.525	0.364	—	—
Thr	6.047	0.978	0.530	0.355	4.65	1.082	0.509	0.336	—	—
Trp	3.540	1.027	0.464	0.332	1.33	0.805	0.501	0.395	—	0
Tyr	1.917	0.549	0.495	0.341	4.52	1.214	0.489	0.355	—	—
Val	8.850	0.919	0.480	0.333	10.90	1.502	0.515	0.380	+	+
Average			0.508	0.349			0.495	0.334	++	++
Standard deviation			0.090	0.101			0.095	0.113		

*—, $p \geq 0.05$.
†0, <10 amino acids in one or both protein classes.
‡+, $0.01 \leq p < 0.05$.
§++, $p < 0.01$.

The PDB codes for the helical membrane proteins analyzed are as follows: 1C3W (0.451), 1DXR (0.472), 1EUL (0.419), 1EZV (0.413), 1F88 (0.439), 1FX8 (0.466), 1J95 (0.424), 1KZU (0.415), 1MSL (0.387), 2OCC (0.451), 1QLA (0.426). The average packing values for the membrane helices are in parentheses. The hydrophobic boundaries of the helices were assigned based on the position of basic and acidic residues that bracketed the central hydrophobic portion of the helix.

The PDB codes for the soluble α-bundle proteins analyzed are as follows: 1A17 (0.385), 1B5L (0.411), 2BCT (0.438), 2CCY (0.393), 1DVK (0.413), 1ECM (0.369), 1FT1 chain A (0.447), 1IHB (0.423), 1LIS (0.389), 1LRV (0.417), 1POC (0.420), 1VDF (0.387). The average helix packing values are in parentheses. The helix assignments for soluble proteins were taken directly from the PDB files. Only helices with nine or more residues were considered.

The PDB codes for the soluble α-bundle domains analyzed are as follows with the average packing values for the helices in parentheses: 1A26 (0.393), 1A5T (0.390), 1BUC (0.429), 1CHK (0.441), 1CIY (0.407), 1DIK (0.452), 1FUP (0.404), 1KNY (0.408), 1MTY (0.430), 1VNS (0.450), 1YGE (0.400). The helix assignments for the soluble protein domains were taken directly from the PDB files. Only helices with nine or more residues were considered.

RESULTS

Helix packing in membrane and soluble α-bundle proteins

The internal packing of membrane proteins of known structure has been studied in detail using the method of occluded surfaces. Table 1 presents a comparison of the average residue packing values of 11 helical membrane proteins and 23 soluble α-bundle proteins and α-bundle domains. The average amino acid packing values were calculated by taking the sum of the packing values for each individual amino acid of a given type (e.g., all Ala residues in membrane proteins) and dividing by the total number of those amino acids. The average protein packing values were calculated by taking the sum of the packing values for each individual amino acid in a helix of the protein and dividing by the total number of amino acids.

We first compare the average packing values for amino acids in helices of membrane proteins (average PV = 0.441) with that for soluble α-bundle proteins and α-bundle domains (average PV = 0.418). The average packing value for membrane proteins is distinctly higher. Based on the number of residues that compose the data set and the standard deviation between the individual amino acid packing values, it is possible to assess the significance of the difference between the average or mean packing values using the *z*-test. The *z*-test indicates that there is a >99.9% probability ($p < 0.001$) that the higher average packing value calculated for membrane proteins is statistically significant. The significance of this result is better appreciated when one considers that 8 of the 11 membrane proteins studied have

higher packing values than the average soluble protein packing value. The membrane protein with the lowest packing value (0.387) is the mechanosensitive channel (1MSL), which has a nonselective central ion pore.

One of the advantages of the OS method is that packing values are calculated for individual amino acids. This allows us to assess how different amino acids contribute to the average protein packing density. Table 1 summarizes the average residue packing values for each of the 20 amino acids. Of note is the observation that glycine has the highest overall packing value in membrane proteins (0.524), followed by proline (0.507) and alanine (0.488). In soluble proteins, Cys (0.497) has the highest overall packing value, followed by Phe (0.491) and Ile (0.482). When comparing the packing value differences between membrane and soluble proteins, the *z*-test (Table 1, *right column*) provides a convenient way to assess significance. For instance, the higher packing values for glycine and proline in membrane proteins are significant, whereas the higher packing value observed for alanine (0.488 vs. 0.472) is not significant.

A second way to assess how different amino acids contribute to the average protein packing value is to calculate the occurrence of each amino acid type as a function of packing value. Fig. 4 lists the most abundant amino acids having high packing values (>0.55), intermediate packing values (0.55–0.30), and low packing values (<0.30). Only amino acids with an occurrence of >5% are included. The most striking result of this analysis is the high abundance of Gly, Ser, and Thr in the tightly packed category for membrane proteins. In contrast, the most tightly packed residues in soluble α-bundle proteins are Ala, Leu, Val, Gly, Ile, and Phe. Four of these residues (Leu, Ala, Ile, and Val) have high occurrences in the "a" and "d" positions of left-handed coiled-coil structures (Cohen and Parry, 1990). Another interesting observation is that phenylalanine, which has a low average residue packing value (0.408) in membrane proteins and a very high packing value in soluble proteins (0.491), contributes to all three packing ranges in membrane proteins. This implies that Phe is a very versatile amino acid in terms of membrane protein structure.

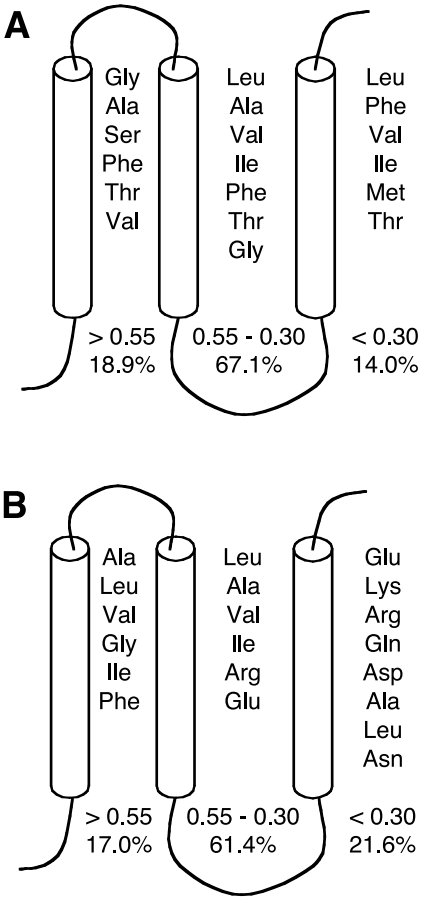

FIGURE 4 The most abundant amino acids in three different packing ranges in membrane (*A*) and soluble α-bundle (*B*) proteins. Packing values were calculated for all amino acids in 11 helical membrane and 23 soluble α-bundle proteins using the OS method. Amino acids with high packing values (>0.55) represent ~20% of all transmembrane residues. The occluded surface analysis indicates that 75% of these residues fit the standard definition of being "buried," i.e., having a solvent-accessible surface area of 20% or less.

Helix-helix interactions in membrane and soluble α-bundle proteins using contact plots

The first challenge in characterizing helix-helix interactions is to define the helix-helix interface and the specific residues that are involved in mediating helix-helix association. The approach we have taken is to generate contact plots for all interacting pairs of helices in a protein structure. The contact plots have several advantages, including that they are easy to generate and there is a straightforward visual correlation between the contact plot and the geometry of the interacting helix pair. The interfacial residues are defined as those located at a local minimum. For each turn of an α-helix there are either one or two residues that fit this definition. In an idealized left-handed coiled coil of helices, our definition would correspond to the residues at the "a" and "d" positions. The other positions are considered noninterfacial. These may be involved in helix-helix, helix-lipid, or helix-water interactions. An alternative definition of interfacial residues involving contact between Voronoi polyhedra that share common edges is less restrictive and would include residues in the "e" and "g" positions of a heptad repeat (Adamian and Liang, 2001).

The contact plots shown in Fig. 3 illustrate that helices have a wide range of relative orientations. The helix interface can be very closely packed in the region where the helices cross and can be very loosely packed in the region where the helices diverge. We have constructed contact plots for 142 interacting helix pairs in membrane proteins and 190 interacting helix pairs in soluble α-bundle proteins and α-bundle domains. Fig. 5 presents the distribution of the

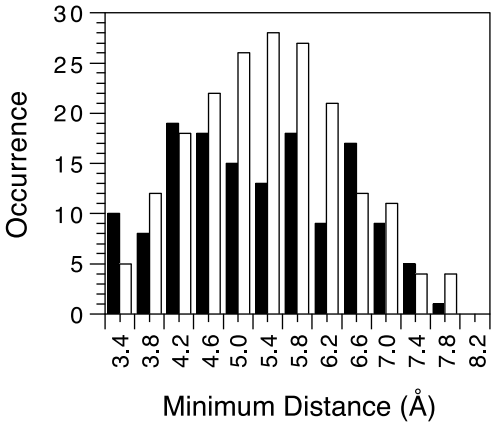

FIGURE 5 Minimal interhelical backbone distances in membrane proteins and in soluble α-bundle proteins. For this analysis, the helices in 11 helical membranes and 23 soluble α-bundle proteins were divided into interacting pairs, and interhelical distances were calculated from the backbone atom coordinates. The distribution of the minimal interhelical distances are summed in 0.4 Å intervals and plotted for membrane proteins (*filled bars*) and for soluble α-bundle proteins (*open bars*). The average minimal interhelical distance is 5.10 Å (standard deviation 1.12 Å, variance 1.26 Å) in membrane proteins and 5.22 Å (standard deviation 0.99 Å, variance 0.98 Å) in soluble α-bundle proteins.

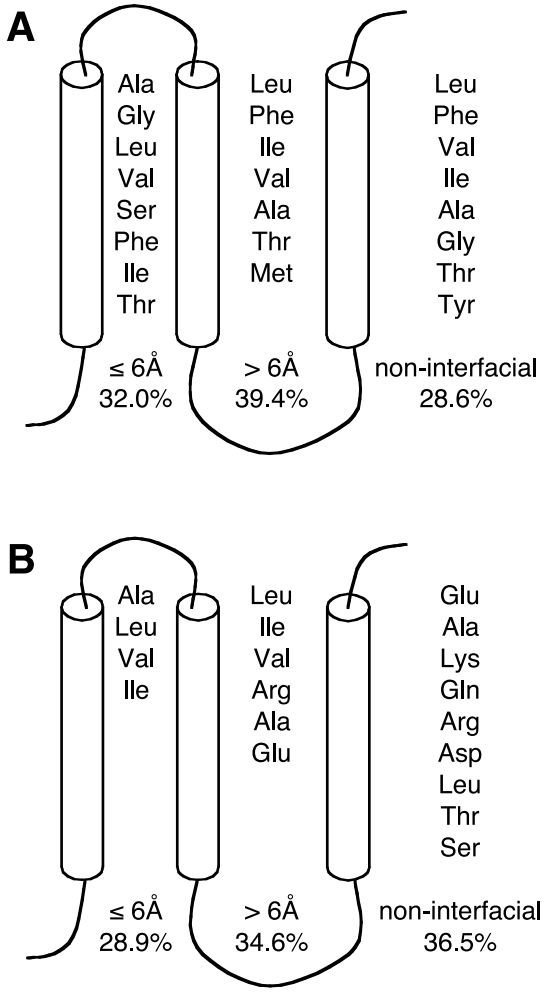

FIGURE 6 Amino acid occurrence in helix interfaces. Only those residues whose occurrence is >5% are listed. The helix interfaces (≤6 Å) of membrane proteins are composed of eight amino acids, which are statistically overrepresented (i.e., >1/20), whereas in soluble proteins only four amino acids are statistically overrepresented. The four dominant amino acids in soluble proteins (Ala, Leu, Val, and Ile) correspond to the most abundant amino acids in the "a" and "d" positions of left-handed coiled-coil proteins (Cohen and Parry, 1990).

minimum backbone-to-backbone distances between helices in membrane proteins (filled bars) and soluble proteins (open bars). The minimum distances range from 3.0 to 7.8 Å. For soluble proteins, the distribution of minimum distances is symmetric about an average backbone-to-backbone distance of 5.22 Å. For membrane proteins, the distribution is clearly different. The higher relative occurrence of short distances is consistent with the high abundance of residues with small side chains in the most tightly packed class (>0.55) in the packing analysis shown in Fig. 4. The average minimum distance for membrane proteins is 5.10 Å. Based on the *z*-test, however, the difference between the average minimum distances for membrane and soluble proteins is not statistically significant ($p \geq 0.05$).

To characterize the nature and distribution of amino acids in interfacial and noninterfacial positions, we divided the 2118 residues in membrane protein helices and the 2604 residues in soluble α-bundle protein helices into three categories: those at a local minimum in the helix contact plots and having a minimum backbone-to-backbone distance of ≤6 Å, those at a local minimum in the helix contact plots and having a minimum backbone-to-backbone distance of >6 Å, and those not at a local minimum. The first category represents ~50% of the interfacial residues, and we use the terms interface and interfacial residues for these amino if not otherwise stated. A 6 Å minimum distance between backbone atoms corresponds to an axial separation of ~11 Å, and includes most residues that would be either close or moderately packed in helix interfaces. (For comparison, the average minimum backbone-to-backbone distance in the long left-handed coiled coil of GCN4 is 5.7 Å, while the average axial separation for membrane and soluble proteins is ~9.6 Å).

Table 2 lists the occurrence of contact residues in helix interfaces by amino acid type for membrane proteins and for soluble α-bundle proteins and α-bundle domains. The most striking result in Table 2 is that in membrane proteins, compared to soluble α-bundle proteins, there are twice as many residues that have occurrences of >5%. This clearly shows that membrane proteins have a more diverse set of interactions mediating close helix-to-helix contacts.

Fig. 6 illustrates the results of the contact plot analysis by listing the most abundant residues with occurrences of 5% or greater in the three categories defined above. In the first category (interfacial residues where the minimum backbone

TABLE 3 Weighted propensities for the amino acids in helical membrane and soluble α-bundle proteins

	Membrane Proteins			Soluble α-Bundle Proteins		
	Interface $\leq$ 6 Å	Interface $>$ 6 Å	Noninterface	Interface $\leq$ 6 Å	Interface $>$ 6 Å	Noninterface
Ala	21.345	5.778	9.037	32.440	4.061	7.045
Arg	0.910	2.401	2.748	2.714	7.705	8.839
Asn	2.658	2.524	0.531	1.680	2.064	7.833
Asp	0.903	3.111	0.092	0.949	2.347	12.625
Cys	2.457	1.125	0.766	4.069	0.999	0.398
Gln	0.461	2.382	0.828	1.667	4.733	10.235
Glu	0.982	5.404	0.136	0.638	7.245	20.035
Gly	19.243	2.583	6.602	2.979	1.514	5.862
His	1.252	6.626	0.531	3.422	3.268	2.137
Ile	4.323	10.341	13.319	8.615	9.694	1.535
Leu	8.815	18.457	18.440	18.528	29.897	2.935
Lys	0.867	2.287	1.691	1.749	3.387	17.587
Met	4.718	5.617	3.593	1.830	4.103	3.133
Phe	5.255	9.867	15.165	3.338	6.371	1.492
Pro	7.372	1.048	0.751	2.041	0.956	2.435
Ser	10.765	3.478	2.600	3.377	2.396	6.512
Thr	5.912	7.015	5.375	5.054	1.730	6.324
Trp	3.636	3.500	3.150	0.798	4.029	0.364
Tyr	1.052	3.452	7.955	5.399	4.516	2.290
Val	8.131	8.601	13.030	14.170	7.726	1.995

separation is $\leq$6 Å), there are eight amino acids in membrane proteins that are statistically overrepresented, whereas in soluble proteins there are only four amino acids. For both membrane and soluble proteins, the most abundant interfacial residue is alanine. Membrane and soluble proteins diverge at the second most abundant amino acid, which is glycine (12.1%, a total of 82) in membrane proteins and leucine (15.8%, a total of 119) in soluble α-bundle proteins (Table 2, Fig. 6).

Finally, a comparison of the results summarized in Table 1 and Fig. 6 shows that in general, small amino acids pack more tightly in membrane proteins, whereas large amino acids pack more tightly in soluble α-bundle proteins. For instance, Leu, Ile, and Val have higher occurrences and higher packing values in soluble proteins, whereas Gly, Ser, and Thr have higher occurrences and higher packing values in membrane proteins.

Amino acid propensities in helix interfaces

The packing and contact analyses, which are summarized in Figs. 4 and 6, are based on total occurrences. Leu and Ala have the highest total occurrences in the helices of both membrane and soluble α-bundle proteins (Table 1), and consequently it is not surprising that these two residues rank high in the analyses. A more complete picture for how membrane and soluble proteins use different amino acids to mediate helix interactions is given by calculating amino acid propensities. Propensities will not be influenced by total occurrences, but rather will provide an "intrinsic" measure of whether an amino acid is likely to be found in an interfacial or noninterfacial position. We define interfacial propensity as the interfacial occurrence of an amino acid (Table 2) divided by its total occurrence (Table 1). Table 2 lists the interfacial propensity by amino acid type for residues with backbone-to-backbone separations of $\leq$6 Å. The amino acids with the highest interfacial propensities are Pro and Cys in membrane and soluble proteins, respectively.

The drawback of interfacial propensities is that it does not take into account the total interfacial occurrence of an amino acid. As a result, residues that are fairly rare in helices (e.g., Cys, with an occurrence of only 2%) can rank very high. The way to combine the intrinsic propensities with total occurrences is to calculate a "weighted propensity" by multiplying the interfacial (or noninterfacial) propensity by the interfacial (or noninterfacial) occurrence. Table 3 lists the amino acids by weighted propensities for both interfacial and noninterfacial amino acids. Fig. 7 depicts the results for those residues with a weighted propensity of $>$5%. This analysis complements the analyses based on total occurrences and on propensities alone. The most striking result from Fig. 7 is the high weighted propensities of the small residues Ala, Gly, and Ser for mediating helix-to-helix interactions in membrane proteins.

Occurrence and packing of amino acids in helices with left- and right-handed crossing angles and with parallel and antiparallel orientations

The first two sections above greatly extend our previous studies using the occluded surface analysis and helix contact plots to characterize amino acid packing and helix-helix interactions, respectively. The results of these

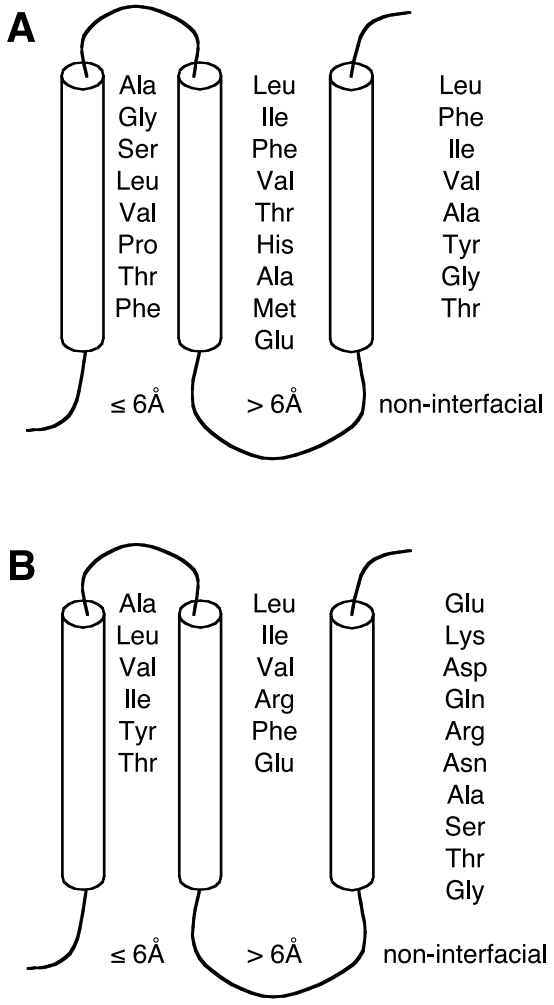

FIGURE 7 Amino acid propensity in helix interfaces. The most abundant residues in helix interfaces are shown for membrane (*A*) and soluble α-bundle (*B*) proteins based on propensity. Only those residues whose occurrence is >5% are listed.

methods can now be combined to investigate the differences in packing for helices having left- or right-handed crossing angles and having parallel or antiparallel helix orientations.

Crick (1953) originally introduced the "knobs-into-holes" model to describe the role of steric contacts and surface complementarity in helix-to-helix packing. He found that the optimal packing angle between helices was at $+20°$, corresponding to the angle formed for helices forming left-handed coiled coils. A second preferred packing angle of $-70°$ was also described. Subsequently, this basic model has been refined (Richmond and Richards, 1978) and several additional models have been proposed (Chothia et al., 1977, 1981; Walther et al., 1996). Most recently, Bowie (1997a) and Walther et al., (1998) have shown that if one accounts for the statistical bias toward crossing angles with perpendicular orientations, there is a preference for helices to be aligned in a parallel or antiparallel fashion. Bowie (1997a) argued that this preference does not agree with the regular packing models. In this section we address the question of whether differences are observed in the nature and distribution of residues in transmembrane helices having left- and right-handed crossing angles. The results are compared to soluble α-bundle proteins.

We determined the packing angles for all of the helix pairs used in the contact analysis. The distribution of angles agrees with the distributions observed by Senes et al. (2001) and others (Bowie, 1997b; Walther et al., 1998). Left-handed crossing angles are preferred in both membrane (61%:39%) and soluble α-bundle proteins (62%:38%), with antiparallel left-handed orientations being favored in membrane proteins (42% of all helix pairs). Table 4 lists the packing values and occurrences for all interfacial residues in helices that have either left- or right-handed crossing angles for both membrane and soluble α-bundle proteins. The relative occurrences of amino acids in the interfaces of helices with left- and right-handed crossed angles are nearly the same for both membrane and soluble α-bundle proteins. The only notable exception is that of Leu in soluble proteins where there is a much higher interfacial occurrence in helices with left-handed crossing angles.

Table 4 also lists the average packing values for the interfacial residues in helices with left- and right-handed crossing angles in membrane and soluble α-bundle proteins. The data address whether there are statistically significant differences between left- and right-handed packing. The average packing value for helices with left-handed crossing angles in membrane proteins (0.518) is significantly higher ($p < 0.01$) than for helices with right-handed crossing angles (0.501). This difference agrees with the smaller average packing angles that are found in helices with left-hand crossing angles because smaller crossing angles will allow the helices to be more closely associated along their entire length. The average packing value for left-handed helices in membrane proteins is also significantly higher than the average packing values for helices in soluble proteins with either left- or right-handed crossing angles. This may simply reflect the fact that transmembrane helices are much longer, on average, than helices in soluble proteins, and exhibit a strong preference for the optimal $+20°$ left-handed crossing angle characteristic of coiled coils. Bowie (1997a) proposed that this strong preference is due to regular helix packing. In left-handed coiled coils the side chains of one helix pack into the "holes" or "grooves" on the opposing helix and greatly restrict the helix crossing angle. In contrast, Bowie (1997a) found that the helix packing angles in soluble proteins were more variable and not well described by regular packing models. This would imply that the side chains would be less tightly packed and there would be no difference in the packing of helices with left- or right-handed crossing angles.

TABLE 4 Amino acid packing values in helical interfaces with left- and right-handed crossing angles

	Left-Handed Crossing Angle				Right-Handed Crossing Angle			
	Membrane Proteins		Soluble α-Bundle Proteins		Membrane Proteins		Soluble α-Bundle Proteins	
	PV*	%†	PV	%	PV	%	PV	%
Ala	0.545	13.707	0.539	20.083	0.514	15.833	0.559	20.667
Arg	0.517	2.124	0.406	3.520	0.488	0.417	0.388	4.333
Asn	0.562	1.737	0.494	2.277	0.494	3.333	0.450	2.667
Asp	0.565	0.772	0.401	1.656	0.486	1.250	0.380	2.000
Cys	0.506	1.544	0.541	1.656	0.517	2.083	0.485	2.667
Gln	0.536	0.772	0.487	2.484	0.419	1.667	0.397	3.000
Glu	0.488	1.158	0.397	1.242	0.538	1.667	0.420	3.000
Gly	0.573	12.162	0.516	3.313	0.544	10.833	0.541	3.333
His	0.595	2.510	0.460	3.313	0.501	1.667	0.448	3.667
Ile	0.494	6.371	0.506	5.590	0.481	7.083	0.489	8.333
Leu	0.468	11.969	0.498	18.219	0.465	10.417	0.501	11.667
Lys	0.506	1.158	0.356	2.899	0.427	0.833	0.348	2.333
Met	0.490	4.633	0.473	2.484	0.495	4.583	0.543	2.000
Phe	0.476	8.301	0.518	3.106	0.491	5.000	0.510	2.333
Pro	0.547	3.475	0.416	1.449	0.515	5.833	0.476	2.333
Ser	0.546	7.336	0.536	5.176	0.542	7.500	0.479	3.000
Thr	0.504	6.371	0.514	4.141	0.520	5.833	0.519	6.667
Trp	0.481	2.703	0.529	1.242	0.436	4.167	0.460	1.333
Tyr	0.513	2.317	0.477	5.176	0.538	2.083	0.500	4.000
Val	0.498	8.880	0.507	10.973	0.468	7.917	0.531	10.667
Average	0.518		0.498		0.501		0.497	

*PV, packing value.
†%, percent occurrence.

We further determined helix packing as a function of the helix orientation in membrane proteins. In our data set of membrane proteins, antiparallel orientations (66%) are favored over parallel orientations. This agrees with the previous observation of Bowie (1997b). We find that helix pairs with antiparallel orientations are more tightly packed than helix pairs with parallel orientations and have an average residue packing value of 0.522 for those amino acids in the $\leq$6 Å interface. In contrast, helix pairs with parallel orientations have an average packing value of 0.500. This difference is significant ($p < 0.001$). If we consider all interfacial residues, the average packing value of helices with antiparallel orientations (0.495) is still significantly higher ($p < 0.001$) than that for helices with parallel orientations (0.478). These results are in agreement with Bowie (1997b) who showed that helix pairs with antiparallel orientations tend to have larger contacting surfaces.

DISCUSSION

Membrane proteins have a higher diversity of residues in helical interfaces than soluble α-bundle proteins

The combined contact-packing analysis reveals that membrane proteins have a much higher diversity of interhelical interactions than soluble proteins, and that there is a high propensity for small and polar residues in closely packed helix interfaces. The contact plots not only reveal which amino acids line the helix-helix interfaces, they also provide information on pairwise interactions. Tables 5 and 6 present the pairwise interactions for amino acids in the helix interfaces of membrane and soluble α-bundle proteins. The pairwise contacts were calculated for helix interfaces with backbone separations of $\leq$6 Å. As might be expected, there is a much broader distribution of pairwise interactions in membrane proteins than in soluble α-bundle proteins. Significant numbers of pairwise contacts are made between Leu, Ala, Val, and Ile in both membrane and soluble proteins, whereas in membrane proteins there are a large number of contacts that are also made by small and polar residues.

The distributions of pairwise contacts were also calculated for amino acids in helix interfaces of membrane and soluble α-bundle proteins with backbone separations $>$6 Å (data not shown). The distributions were considerably different than those in Tables 5 and 6 and in both cases were dominated by large residues (mainly Leu, Ile, Phe, and Val). The highest pairwise interactions in membrane proteins were Leu-Val (32), Leu-Phe (28), and Leu-Ile (27), while in soluble proteins Leu-Leu (62) contacts dominated the pairwise interactions, followed by Leu-Ile (39) and Leu-Ala (34). This analysis complements the recent work of Adamian and Liang (2001) who investi-

TABLE 5 Pairwise interactions between amino acids in membrane proteins in the ≤6 Å helical interface

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	2	1	3	2	3	1	1	7	2	7	5	0	7	4	6	5	5	2	1	6
Arg	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0
Asn	4	0	1	1	0	0	1	0	0	0	1	0	0	3	0	1	2	0	1	2
Asp	3	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
Cys	3	0	0	0	0	0	0	5	0	1	2	0	0	1	0	0	0	0	0	1
Gln	2	0	0	0	0	0	1	0	1	0	1	0	0	0	0	2	0	0	0	0
Glu	2	0	1	1	0	0	0	1	0	0	0	1	0	1	0	3	3	0	0	0
Gly	6	1	1	0	3	0	2	16	5	4	10	0	3	9	1	5	2	6	2	2
His	2	0	0	0	0	2	0	6	0	1	2	0	0	0	0	1	3	1	0	3
Ile	6	1	0	1	1	0	0	2	1	6	6	0	0	0	3	2	1	0	1	2
Leu	14	0	3	1	1	0	0	10	0	5	9	3	4	3	2	5	1	2	2	4
Lys	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	1	0	0	0
Met	9	0	0	0	0	0	0	2	0	2	1	1	1	7	3	2	1	2	0	0
Phe	5	1	2	0	2	0	0	6	1	2	6	0	5	5	2	5	3	2	1	6
Pro	6	1	0	1	1	0	0	2	0	2	2	0	4	1	3	1	0	1	3	3
Ser	7	2	1	0	0	1	3	5	1	3	7	1	4	3	2	8	5	0	1	5
Thr	3	0	1	0	0	0	0	2	1	3	5	1	2	5	0	4	5	3	0	5
Trp	2	1	0	0	1	1	0	6	0	1	2	0	1	2	1	0	3	4	0	2
Tyr	2	1	1	0	0	0	0	0	0	2	2	0	0	1	2	0	0	0	0	2
Val	5	0	0	0	0	0	0	10	0	2	5	0	0	3	3	4	5	1	1	7

gated the pairwise interactions of residues in membrane proteins using Voronoi constructions to define interacting neighbors. They reached a similar conclusion that membrane proteins exhibit a high diversity of helix interactions, and noted that transmembrane helices have a larger variety of polar-polar interactions than soluble proteins.

The high diversity of helix-helix interactions in membrane proteins is likely to be related to membrane protein function and structure. Unlike soluble proteins, where the functional sites are on the protein surface in active site clefts or grooves, the functional sites in membrane proteins are often in the protein interior. These sites typically contain highly polar amino acids. For instance, Lys-216 in the interior of bacteriorhodopsin is the site of attachment for the protein's retinal chromophore. The Lys-216 side chain is packed against Pro-50 and Ala-53 in the interface between helices 2 and 7 (see Fig. 3 *B*). Lysine, along with Asp, Glu, Arg, Asn, and Gln, are relatively rare in transmembrane helices. The importance of these highly polar amino acids in forming diverse pairwise interactions is not reflected in Tables 5 and 6, which list absolute occurrences. These residues do have high pairwise propensities as seen in the analysis of Adamian and Liang (2001).

TABLE 6 Pairwise interactions between amino acids in soluble α-bundle proteins in the ≤6 Å helical interface

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	42	9	2	1	2	5	2	6	6	9	31	4	5	4	5	7	5	4	9	24
Arg	7	0	1	0	0	1	0	0	0	1	6	0	0	0	1	1	0	0	2	5
Asn	1	1	2	0	0	0	0	0	1	0	4	0	0	1	0	4	2	0	0	0
Asp	2	0	0	0	0	1	0	2	1	0	1	1	0	0	0	1	3	0	2	1
Cys	4	0	0	0	2	0	0	0	3	0	5	0	0	2	0	0	0	1	0	2
Gln	5	2	0	0	0	0	2	1	0	0	3	1	1	2	0	0	0	0	1	3
Glu	1	0	0	0	0	1	0	0	2	2	2	1	0	0	0	0	1	0	0	0
Gly	8	0	0	2	0	1	1	4	1	2	4	1	0	1	0	2	0	0	0	3
His	6	0	1	1	3	0	3	1	0	3	0	1	0	1	0	1	3	0	1	4
Ile	6	2	0	0	0	0	2	1	1	4	9	0	0	1	1	1	2	0	1	7
Leu	26	6	4	1	3	4	3	3	0	13	15	7	6	4	3	8	4	2	5	13
Lys	2	2	0	1	0	1	0	0	1	0	6	0	1	0	0	0	2	0	0	3
Met	3	0	0	1	0	0	0	0	0	1	3	0	0	1	0	0	2	0	0	1
Phe	3	0	2	0	3	2	0	1	1	2	3	0	1	0	1	0	0	0	2	1
Pro	5	1	1	1	0	0	1	0	1	3	2	0	0	1	0	0	1	0	0	0
Ser	5	0	4	1	0	0	0	2	1	2	7	0	0	0	0	4	3	0	1	1
Thr	6	0	2	3	0	1	0	0	2	2	3	2	3	0	2	3	3	0	0	9
Trp	4	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1
Tyr	8	2	0	1	0	1	0	0	2	1	5	0	0	4	0	1	0	0	9	0
Val	15	5	0	1	2	3	1	5	4	7	12	3	1	0	1	1	9	3	4	7

Small residues have high propensities for packing in helix interfaces in membrane proteins

The combined contact-packing analysis shows that small and polar residues have a high weighted propensity to occur in transmembrane helix interfaces and are among the most tightly packed amino acids in membrane proteins. Our analysis strongly argues that helix interactions are both qualitatively and quantitatively different between membrane and soluble proteins. The comparison can be further quantified by plotting the propensity as a function of amino acid volume (Fig. 8, *A* and *B*) and hydrophobicity (Fig. 8, *C* and *D*). For membrane proteins (Fig. 8 *A*), there is a rough linear correlation (R = 0.70) between the propensity and volume. This correlation does not hold for soluble α-bundle proteins (Fig. 8 *B*) (R = 0.28). The observation that small residues have a high propensity for lining the interfaces between helices also agrees with an analysis of surface roughness in helical membrane proteins that shows that in general the lipid-exposed surface is rough, whereas the helix-helix interfaces are smooth (Renthal, 1999).

In contrast, for soluble α-bundle proteins and domains (Fig. 8 *D*), there is a rough linear correlation (R = 0.66) between the weighted interfacial propensity and hydrophobicity. This correlation does not hold for membrane proteins (Fig. 8 *C*) (R = 0.31). This is not surprising because the folding of soluble proteins is driven by the hydrophobic effect. In membrane proteins, large hydrophobic residues (Leu, Phe, Val, and Ile) have the highest occurrences on the lipid-exposed surface of membrane proteins (Fig. 6).

The fact that small and polar residues emerge from the analysis of helix interactions in membrane proteins based on both absolute occurrences and propensities emphasizes their importance. Small and polar residues also occur in the interfaces of soluble proteins; however, neither their absolute occurrences nor interfacial propensities are as striking as in membrane proteins. There are several studies where small/polar residue packing is highlighted. Richmond and Richards (1978), in a comprehensive analysis of helix packing in sperm whale myoglobin, described the packing of Gly-25 and Gly-59 at the crossing point of helices B and E. They suggested that the crossing angle between helices would be inversely correlated with the volume of the central residue in a helix. Reddy and Blundell (1993) showed that the axial separation between helices is dependent on the volume of the interfacial residues. Efimov (1979) showed that the axial separation was less in helices that were packed in a "polar" fashion, i.e., where hydrophilic residues lie on one face of a pair of interacting helices. More recently, Walther et al. (1996) found that helices with different axial separations used different packing patterns. They concluded that Ala has the highest packing flexibility, in agreement with our results where Ala is found to have a high occurrence and propensity in the interfaces of both membrane and soluble α-bundle proteins.

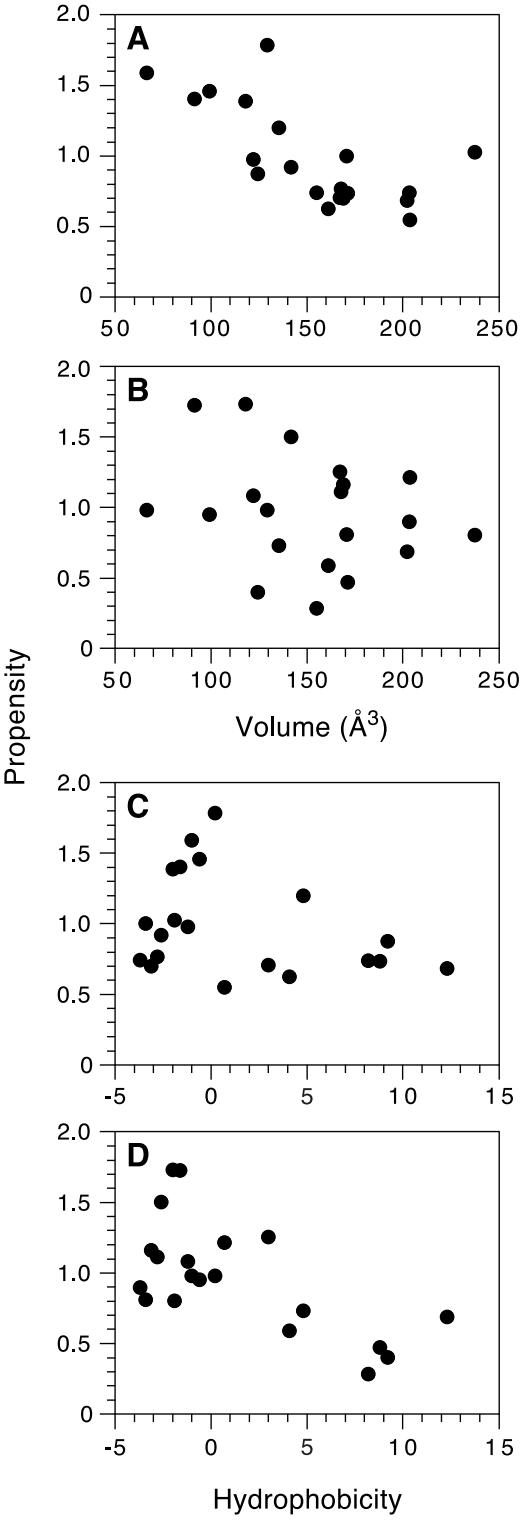

FIGURE 8 Amino acid propensity in helix interfaces as a function of residue volume (*A*, *B*) and hydrophobicity (*C*, *D*). The amino acid volumes were taken from Chothia (1975) and are G, A, S, C, T, P, D, N, V, R, E, Q, H, L, I, M, K, F, Y, W with increasing volume. The hydrophobicity scale proposed by Engelman, et al. (1986) was used for (*C* and *D*), giving the following order with increasing hydrophobicity: R, D, K, E, N, Q, H, Y, P, S, G, T, A, W, C, V, L, I, M, F. For this analysis, a backbone-to-backbone distance cutoff of 6.0 Å was used.

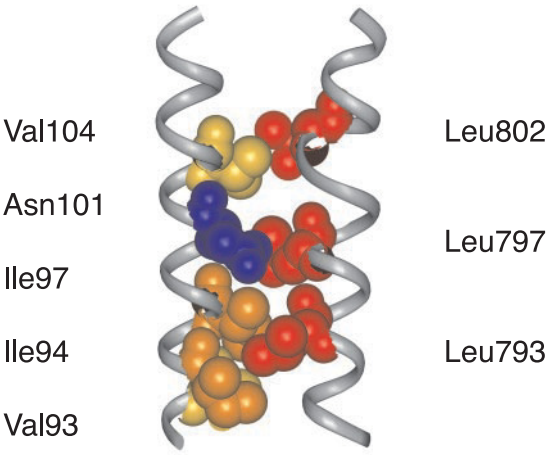

FIGURE 9 Helix-helix interactions between helices M2 and M6 of the Ca^{2+}-ATPase (1EUL). Leu are colored in red, Ile in orange, Val in light orange, and Asn in blue.

Two general motifs exist for helix-helix interactions in membrane proteins

One of the conclusions that emerges from our analysis is that there are statistically significant differences between membrane and soluble α-bundle proteins. The higher diversity of membrane protein interactions and the propensity of small and polar residues in tightly packed interfaces suggests that membrane proteins have at least two general motifs for mediating helix interactions. Both membrane and soluble proteins exhibit "knobs-into-holes" packing exemplified by "leucine zippers" (Cohen and Parry, 1990; Langosch and Heringa, 1998). Membrane proteins, however, have a second general motif, exemplified by the dimer interface of glycophorin A (Lemmon et al., 1992; Smith et al., 2001), in which small and polar residues form smooth surfaces that allow very close approach of the backbones of interacting helices. In this section we present examples of these two general motifs in polytopic membrane proteins and discuss recent studies in which these motifs were found to mediate the dimerization of membrane proteins with single transmembrane helices.

The first motif, which is common to both membrane and soluble proteins, is exemplified by the heptad repeat of leucine residues, LxxLxxxLxx, characteristic of leucine zippers. Analysis of left-handed coiled coils shows that the predominant residues in the "a" and "d" positions of this motif are Leu (33%), Ala (16%), Ile (10%), and Val (7%) (Cohen and Parry, 1990). These four amino acids dominate the core residues involved in helix-helix interactions in soluble proteins (53%, Table 2), and contribute significantly to helix interactions in membrane proteins (42%, Table 2). In an analysis of three membrane proteins, Langosch and Heringa (1998) found that transmembrane helices exhibited knobs-into-holes packing characteristic of left-handed coiled coils in soluble proteins. They concluded that helix packing is less compact than in soluble proteins. This correlates with our results summarized in Fig. 4, which shows that

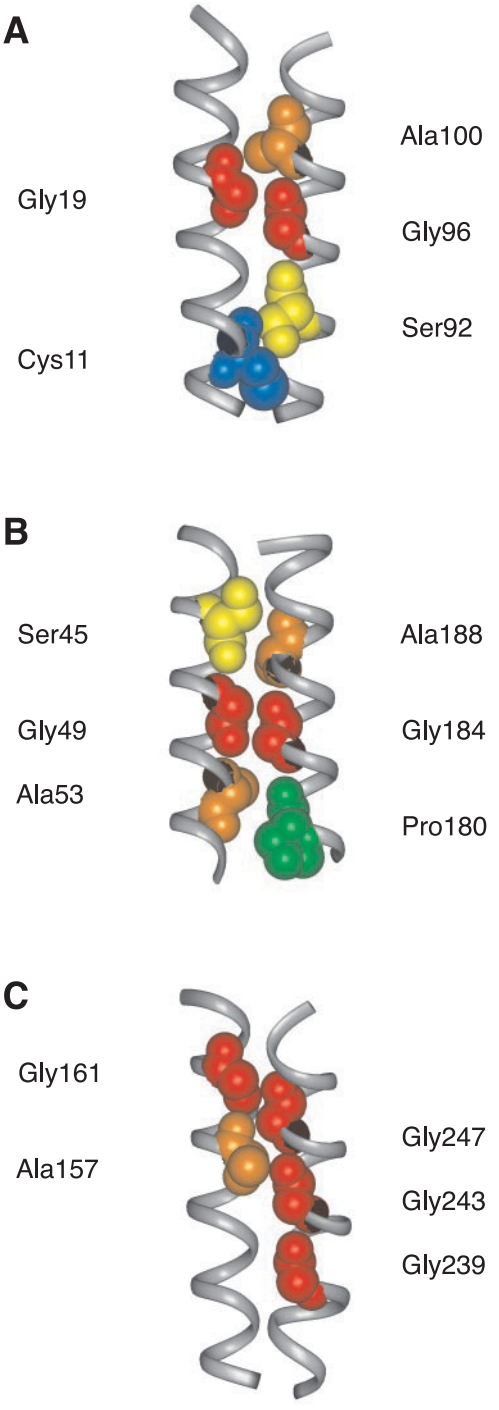

FIGURE 10 Helix-helix interactions in the glycerol facilitator channel (1FX8). (*A*) Helix pair M1 and M4. (*B*) Helix pair M2 and M6. (*C*) Helix pair M5 and M8. The minimum backbone-to-backbone distances for the helix pairs in panels *A–C* are 3.34 Å, 3.08 Å, and 2.98 Å, respectively. Gly are colored in red, Ala in orange, Ser in light yellow, and Pro in green.

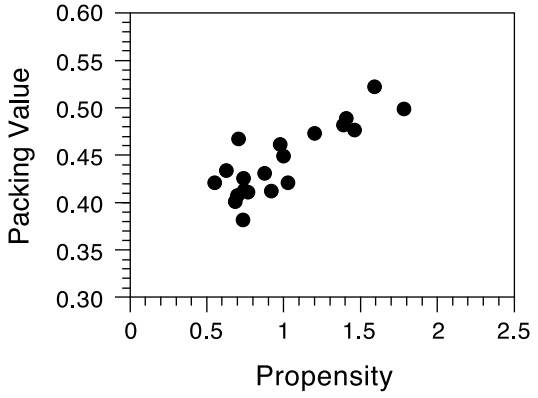

FIGURE 11 Average amino acid packing values as a function of the interhelical propensity. The packing values are taken from Table 1. The interhelical propensities are taken from Table 2 and correspond to the amino acids with backbone-to-backbone separations of ≤6.0 Å.

Leu, Val, Ala, and Ile are the most abundant amino acids in the intermediate packing region (0.55–0.30) in membrane proteins, whereas they dominate the high packing region in soluble proteins. Moreover, Langosch and coworkers showed that a heptad motif of leucine residues can drive the association of designed membrane proteins with single transmembrane helices (Gurezka et al., 1999).

Fig. 9 presents an example of knobs-into-holes packing characteristic of leucine zippers. The helix pair is from the structure of the Ca^{2+}-ATPase (1EUL). The interfacial residues shown are Val-93, Ile-94, Ile-97, Asn-101, and Val-104 for helix M2, and Leu-793, Leu-797, and Leu-802 for helix M6. There are several interesting features exhibited by this helix pair. First, there is an Asn in helix M2 that is involved in interhelical hydrogen bonding. Asn hydrogen bonding is a hallmark of the leucine zipper coiled coil of GCN4. Second, there is a π-bulge near Leu-802 in helix M6. The presence of this distortion in the helix does not disrupt the knobs-into-holes packing arrangement. Finally, the minimum backbone-to-backbone distance is 6.28 Å. As a result, this helix pair falls into the >6 Å category defined for our contact analysis. In this category, Leu and Ile have the highest weighted propensities (Fig. 7).

The second motif that appears to be common in helix-to-helix association in membrane proteins is exemplified by the GxxxG motif observed in glycophorin A. The GxxxG sequence was the most dominant motif found in a statistical analysis of membrane protein sequences (Senes et al., 2000) and dimerization-dependent screens (Russ and Engelman, 2000) for helix interactions in membrane proteins having single membrane spanning helices. In these studies, serine is the most common residue (after glycine) found at the position of one of the glycines in the motif. Our analysis strongly suggests that a similar general motif exists in polytopic membrane proteins where the interfacial positions are occupied by Gly, Ser, Ala, and Thr.

Fig. 10 presents an example of the "small and polar residue" motif from the glycerol-facilitator channel (1FX8). In this protein, there are six full-length transmembrane helices and two "half" helices (Fu et al., 2000). The three helix pairs shown in panels A–C correspond to the six full-length transmembrane helices. The interfaces of all three helix pairs are lined by small and polar residues, and have close (≤6 Å) backbone contacts. Close glycine-glycine contacts are involved in each helix pair, and at least one interfacial glycine is highly conserved across the large family of membrane channels known as aquaporins, of which the glycerol-facilitator channel is a member (Fu et al., 2000). Two of the helix pairs (helices M1–M4 and M2–M6) involve a Ser residue that is in a position to form an interhelical hydrogen bond. Ser-92 on helix M4 may hydrogen-bond to Cys-11 on helix M1, and Ser-45 on helix M2 is in a position to hydrogen-bond to the backbone carbonyl of Ala-192 on helix M6.

A unique structural role for small residues in helix interfaces may involve stabilization of helix dimers via dipolar interactions involving backbone amide C═O and N—H groups or direct hydrogen-bonding interactions involving the polar side chains of Ser, Thr, or Cys. Both types of interactions are facilitated by short interhelical spacing in

TABLE 7 Membrane proteins analyzed

PDB Code	PV	Resolution (Å)	R Factor	Description
1C3W	0.451	1.55	15.8	Bacteriorhodopsin—*Halobacterium salinarium*
1DXR	0.472	2.0	19.4	Photosynthetic Reaction Center—*Rhodopseudomonas viridis*
1EUL	0.419	2.6	25.0	Calcium ATPase—*Oryctolagus cuniculus*
1EZV	0.413	2.3	22.2	Cytochrome bc1 Complex—*Saccharomyces cerevisiae*
1F88	0.439	2.8	18.6	Rhodopsin—*Bos taurus*
1FX8	0.466	2.2	19.7	Glycerol Facilitator (Glpf)—*Escherichia coli*
1J95	0.424	2.8	29.8	Potassium Channel—*Streptomyces lividans*
1KZU	0.415	2.5	22.7	Light Harvesting Complex—*Rhodopseudomonas acidophila*
1MSL	0.387	3.5	26.0	Mechanosensitive Ion Channel—*Mycobacterium tuberculosis*
2OCC	0.451	2.3	20.9	Cytochrome C Oxidase—*Bos taurus*
1QLA	0.426	2.2	21.2	Fumarate Reductase—*Wolinella succinogenes*

TABLE 8 Soluble α-bundle proteins and domains analyzed

PDB Code	PV	Resolution (Å)	R Factor	Description
1A17	0.385	2.45	20.1	Ser/Thr Protein Phosphatase 5—*Homo sapiens*
1A26	0.393	2.25	16.8	Poly (ADP-Ribose) Polymerase—*Gallus gallus*
1A5T	0.390	2.2	20.5	Clamp-Loading Complex of DNA Polymerase III—*Escherichia coli*
1B5L	0.411	2.1	21.4	Ovine Interferon Tau—*Pichia pastoris*
2BCT	0.438	2.9	21.1	Murine—Catenin—*Mus musculus*
1BUC	0.429	2.5	19.3	Butyryl-CoA Dehydro fromgenase—*Megasphaera elsdenii*
2CCY	0.393	1.67	18.8	Ferricytochrome C′—*Rhodospirillum molischianum*
1CHK	0.441	2.4	18.1	Chitosanase—*Streptomyces sp.*
1CIY	0.407	2.25	16.3	CrylA(a) insecticidal toxin—*Bacillus thuringiensis*
1DIK	0.452	2.3	18.2	Pyruvate Phosphate Dikinase—*Clostridium symbiosum*
1DVK	0.413	2.15	20.3	Splicing Factor Prp18—*Saccharomyces cerevisiae*
1ECM	0.369	2.2	19.2	Chorismate Mutase—*Escherichia coli*
1FT1A	0.447	2.25	21.0	Protein Famesyltransferase—*Rattus norvegicus*
1FUP	0.404	2.0	18.5	Fumarase C—*Escherichia coli*
1IHB	0.423	1.95	20.9	Cyclin-Dependent Kinase 6 Inhibitor—*Homo sapiens*
1KNY	0.408	2.5	16.8	Kanamycin Nucleotidyltransferase—*Staphylococcus aureus*
1LIS	0.389	1.9	18.7	Lysin
1LRV	0.417	2.6	20.4	Leucine-Rich Repeat Variant—*Azotobacter vinelandii*
1MTY	0.430	1.7	18.3	Methane Monooxygenase Hydroxylase—*Methylococcus capsulatus*
1POC	0.420	2.0	19.2	Phospholipase A2
1VDF	0.387	2.05	17.6	Cartilage Oligomeric Matrix Protein—*Rattus norvegicus*
1VNS	0.450	1.66	18.0	Vanadium Chloroperoxidase—*Curvularia inaequalis*
1YGE	0.400	1.4	19.7	Lipoxygenase-1

the region of Gly-Gly and Gly-Ala contacts. The importance of hydrogen bonding interactions in hydrophobic protein interiors cannot be overstated because hydrogen bond strengths are much higher in membrane environments (Pace, 2000).

We have examined all potential hydrogen bonding interactions of transmembrane Ser and Thr residues by looking at the distances between the side chain hydroxyl oxygens and all other heteroatoms within 3.4 Å. Based on this analysis (data not shown), most Ser (70%) and Thr (79%) residues in transmembrane helices hydrogen bond to the i-3 or i-4 backbone carbonyl. Nevertheless, ~20–30% of Ser and Thr residues in membrane proteins are in a position to form interhelical hydrogen bonds. Of the Ser and Thr residues that can form interhelical hydrogen bonds to backbone C═O and N—H groups across the helix interface, the hydrogen-bonding partners are predominantly Ser (28%) and Ala (16%), similar to that seen in the interface of the glycerol facilitator protein.

It is interesting to note in regard to interhelical hydrogen bonding that Ser and Thr were largely ineffective for driving helix association in the context of model transmembrane helices containing predominantly leucine (Gratkowski et al., 2001; Zhou et al., 2001). In these studies, the more polar (and much less abundant) amino acids, such as Asn, Asp, Glu, and Gln, were able to drive dimerization. We suggest that Ser and Thr are ineffective in the context of large bulky residues, and that the "small and polar" motif provides specificity for dimer formation in both single-pass and polytopic membrane proteins.

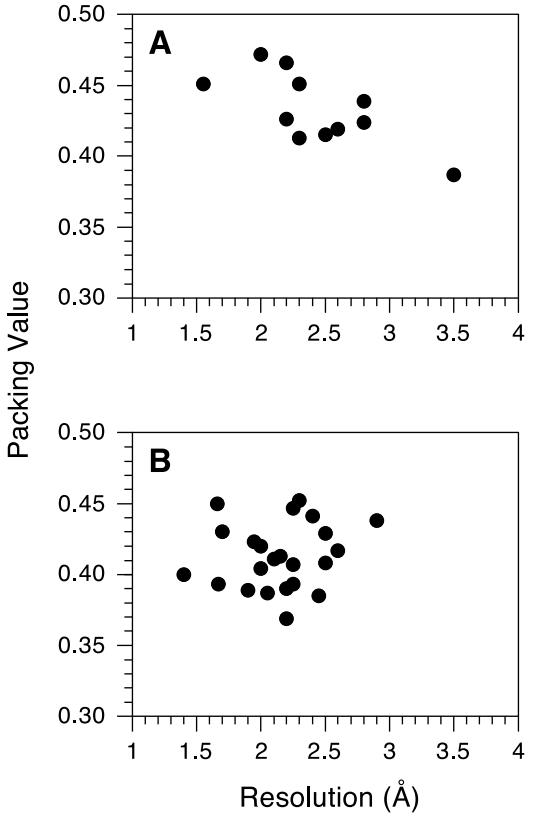

FIGURE 12 Average amino acid packing values as a function of the resolution for membrane (*A*) and soluble proteins (*B*). The packing values are taken from Table 1. The resolutions are taken from Table 7 for membrane proteins and from Table 8 for soluble proteins.

A strong correlation exists between helix packing and interhelical propensity

In soluble proteins, protein stability is closely correlated with the packing of core residues (Richards, 1997). For instance, increased packing appears to be one mechanism by which the extremely stable hyperthermophilic proteins gain increased stability over their mesophilic counterparts (DeDecker et al., 1996). One of the motivations for the current study was to combine the analysis of amino acid packing in membrane proteins with an analysis of helix contacts to address the mechanism of membrane protein stability.

The results in Table 2 indicate that the high packing values in membrane proteins are associated with interfacial interactions. The correlation between packing and interfacial interactions can be further quantified by plotting the packing value as a function of the propensity to occur within the helical interface (Fig. 11). A linear correlation ($R = 0.84$) exists between the packing values determined by the OS method and the interfacial propensity derived from the helix contact plots. These data strongly argue that membrane protein stability (as expressed by packing) has a strong contribution from small and polar interfacial amino acids.

Finally, the correlation between packing and interfacial propensity suggests that interfacial propensities may provide a useful scale for predicting the relative orientation of transmembrane helices. Helical faces with a large number of small and polar residues would be predicted to be closely packed in helix-helix interfaces. Our analysis agrees with that of Rees et al. (1989) that the amino acids on the lipid-exposed surface of membrane proteins are more hydrophobic than interior residues. Using the hydrophobicity scale proposed by Engelman et al. (1986), the residues in the helix interfaces of membrane proteins are less hydrophobic on average (-1.07) than the noninterfacial residues (-1.42). The data also agree with the conclusions of Stevens and Arkin (1999) that the hydrophilic moment of the transmembrane helices is often not oriented toward the center of the transmembrane helical bundle. Rather, the helical hydrophilic moment is often oriented between tightly packed helix pairs containing polar residues, as shown in Figs. 9 and 10. Together, these data illustrate a high degree of complexity in the internal architecture of membrane proteins (compared to soluble helical proteins) and reveal basic strategies used by membrane proteins for forming tight interactions between hydrophobic helices in membrane environments.

APPENDIX

Tables 7 and 8 list the packing values and resolution for the 11 helical membrane proteins and 23 helical α-bundle proteins and protein domains used in our analysis. For the 23 soluble proteins, the average resolution was 2.15 Å (standard deviation 0.34 Å and average R factor of 19.1). For the 11 membrane proteins, the average resolution was 2.43 Å (standard deviation 0.51 Å and an average R factor of 21.9). Fig. 12 plots the packing values of the membrane (*A*) and soluble proteins (*B*) used in our analysis as a function of their resolution. For the membrane protein structures, the packing values tend to be higher for the higher-resolution structures. The six highest-resolution structures have packing values of 0.451 (1C3W, 1.55 Å), 0.472 (1DXR, 2.0 Å), 0.466 (1FX8, 2.2 Å), 0.426 (1QLA, 2.2 Å), 0.413 (1EZV, 2.3 Å), and 0.451 (2OCC, 2.3 Å). The packing values for these structures with the exception of 1EZV are all well above the average packing value (0.418) of the 23 soluble α-bundle proteins.

We thank Erwin London and Stuart McLaughlin for valuable discussions.

This work was supported in part by National Institutes of Health Grant GM-46732 (to S.O.S.).

REFERENCES

Adamian, L. J., and J. Liang. 2001. Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* 311:891–907.

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.

Chothia, C. 1975. Structural invariants in protein folding. *Nature.* 254: 304–308.

Chothia, C., M. Levitt, and D. Richardson. 1977. Structure of proteins: packing of alpha-helices and pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* 74:4130–4134.

Chothia, C., M. Levitt, and D. Richardson. 1981. Helix to helix packing in proteins. *J. Mol. Biol.* 145:215–250.

Cohen, C., and D. A. Parry. 1990. Alpha-helical coiled coils and bundles: how to design an alpha-helical protein. *Proteins.* 7:1–15.

Crick, F. H. C. 1953. The packing of α-helices: simple coiled-coils. *Acta Crystallogr.* 6:689–697.

DeDecker, B. S., R. O'Brien, P. J. Fleming, J. H. Geiger, S. P. Jackson, and P. B. Sigler. 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol.* 264:1072–1084.

Efimov, A. V. 1979. Packing of alpha-helices in globular-proteins: layer-structure of globin hydrophobic cores. *J. Mol. Biol.* 134:23–40.

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. U.S.A.* 97:5796–5801.

Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179:125–142.

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.

Engelman, D. M., and G. Zaccai. 1980. Bacteriorhodopsin is an inside-out protein. *Proc. Natl. Acad. Sci. U.S.A.* 77:5894–5898.

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. X., 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.

Gratkowski, H., J. D. Lear, and W. F. DeGrado. 2001. Polar side chains drive the association of model transmembrane peptides. *Proc. Natl. Acad. Sci. U.S.A.* 98:880–885.

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

Javadpour, M. M., M. Eilers, M. Groesbeek, and S. O. Smith. 1999. Helix packing in polytopic membrane proteins: role of glycine in transmembrane helix association. *Biophys. J.* 77:1609–1618.

Langosch, D., and J. Heringa. 1998. Interaction of transmembrane helices by a knobs-into-holes packing characteristic of soluble coiled coils. *Proteins.* 31:150–159.

Lemmon, M. A., J. M. Flanagan, H. R. Treutlein, J. Zhang, and D. M. Engelman. 1992. Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry.* 31:12719–12725.

Michie, A. D., C. A. Orengo, and J. M. Thornton. 1996. Analysis of domain structural class using an automated class assignment protocol. *J. Mol. Biol.* 262:168–185.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. 1997. CATH: a hierarchic classification of protein domain structures. *Structure.* 5:1093–1108.

Pace, C. N. 2000. Polar group burial contributes more to protein stability than nonpolar group burial. *Biochemistry.* 40:310–313.

Pattabiraman, N., K. B. Ward, and P. J. Fleming. 1995. Occluded molecular surface: analysis of protein packing. *J. Mol. Recognit.* 8:334–344.

Reddy, B. V. B., and T. L. Blundell. 1993. Packing of secondary structural elements in proteins. Analysis and prediction of inter-helix distances. *J. Mol. Biol.* 233:464–479.

Rees, D. C., L. DeAntonio, and D. Eisenberg. 1989. Hydrophobic organization of membrane proteins. *Science.* 245:510–513.

Renthal, R. 1999. Transmembrane and water-soluble helix bundles display reverse patterns of surface roughness. *Biochem. Biophys. Res. Commun.* 263:714–717.

Richards, F. M. 1997. Protein stability: still an unsolved problem. *Cell. Mol. Life Sci.* 53:790–802.

Richards, F. M., and C. E. Kundrot. 1988. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins.* 3:71–84.

Richards, F. M., and W. A. Lim. 1993. An analysis of packing in the protein-folding problem. *Q. Rev. Biophys.* 26:423–498.

Richmond, T. J., and F. M. Richards. 1978. Packing of alpha-helices: geometrical constraints and contact areas. *J. Mol. Biol.* 119:537–555.

Russ, W. P., and D. M. Engelman. 1999. TOXCAT: a measure of transmembrane helix association in a biological membrane. *Proc. Natl. Acad. Sci. U.S.A.* 96:863–868.

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.

Senes, A., I. Ubarretxena-Belandia, and D. M. Engelman. 2001. The Cα-H· · ·O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. U.S.A.* 98: 9056–9061.

Smith, S. O., D. Song, S. Shekar, M. Groesbeek, M. Ziliox, and S. Aimoto. 2001. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry.* 40:6553–6558.

Stevens, T. J., and I. T. Arkin. 1999. Are membrane proteins "inside-out" proteins? *Proteins.* 36:135–143.

Ulmschneider, M. B., and M. S. P. Sansom. 2001. Amino acid distributions in integral membrane protein structures. *Biochim. Biophys. Acta.* 1512: 1–14.

Walther, D., F. Eisenhaber, and P. Argos. 1996. Principles of helix-helix packing in proteins: the helical lattice superposition model. *J. Mol. Biol.* 255:536–553.

Walther, D., C. Springer, and F. E. Cohen. 1998. Helix-helix packing angle preferences for finite helix axes. *Proteins.* 33:457–459.

Zhou, F. X., H. J. Merianos, A. T. Brunger, and D. M. Engelman. 2001. Polar residues drive association of polyleucine transmembrane helices. *Proc. Natl. Acad. Sci. U.S.A.* 98:2250–2255.