Helix Packing in Polytopic Membrane Proteins: Role of Glycine in Transmembrane Helix Association

Maryam M. Javadpour,* Markus Eilers,*# Michel Groesbeek,* and Steven O. Smith#
*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114; and #Department of Biochemistry and Cell Biology, State University of New York at Stony Brook, Stony Brook, New York 11794-5215 USA

ABSTRACT The nature and distribution of amino acids in the helix interfaces of four polytopic membrane proteins (cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center of *Rhodobacter sphaeroides*, and the potassium channel of *Streptomyces lividans*) are studied to address the role of glycine in transmembrane helix packing. In contrast to soluble proteins where glycine is a noted helix breaker, the backbone dihedral angles of glycine in transmembrane helices largely fall in the standard α-helical region of a Ramachandran plot. An analysis of helix packing reveals that glycine residues in the transmembrane region of these proteins are predominantly oriented toward helix-helix interfaces and have a high occurrence at helix crossing points. Moreover, packing voids are generally not formed at the position of glycine in folded protein structures. This suggests that transmembrane glycine residues mediate helix-helix interactions in polytopic membrane proteins in a fashion similar to that seen in oligomers of membrane proteins with single membrane-spanning helices. The picture that emerges is one where glycine residues serve as molecular notches for orienting multiple helices in a folded protein complex.

INTRODUCTION

Glycine is unique among the amino acids in its lack of a side chain. In soluble proteins, glycine residues are often found in loop regions and β-turns because of the absence of steric interactions and the entropic cost of tethering glycine in helical secondary structure. In fact, glycine is generally known to be a "helix-breaker" and ranks with proline in most measurements of helical propensities (O'Neil and DeGrado, 1990). Nevertheless, glycine occurs frequently in the transmembrane helices of membrane proteins (Landolt-Marticorena et al., 1993). The presence of glycine in hydrophobic membrane sequences is allowed as a result of strong hydrogen bonding interactions of the backbone amides and carbonyls that overwhelm the entropic costs of constraining the glycine backbone. The high frequency of glycine residues in transmembrane helices suggests a structural role, which is distinct from that in soluble proteins as suggested by studies on model peptides (Li and Deber, 1992).

The structural role of glycine residues in single-pass membrane proteins has been well studied. In several membrane proteins having only a single membrane-spanning helix, glycine is involved in protein dimerization through specific packing interactions (Lemmon and Engelman, 1994). For example, in glycophorin A, the major glycoprotein in erythrocyte cell membranes, Gly-79 and Gly-83 have been shown to contribute to a dimerization motif consisting of seven critical interfacial residues (Lemmon et al., 1992a, b, 1994). Substitution of either Gly-79 or Gly-83 with larger hydrophobic amino acids was shown to disrupt dimerization. NMR measurements revealed that the two transmembrane glycine residues allowed favorable van der Waals contacts with valine side chains on the opposing helix (Smith and Bormann, 1995; MacKenzie et al., 1997). In the *neu/erbB-2* receptor tyrosine kinase a single valine-to-glutamic acid substitution in the transmembrane domain is responsible for constitutive activation of the receptor. This glutamic acid substitution is only effective in the context of a three-residue sequence, Val-663–Glu-664–Gly-665 (Burke et al., 1997). Substitution of Gly-665 by valine blocks the transforming activity of the V664E mutation (Burke et al., 1997). Magic angle spinning NMR studies have recently shown that Gly-665 lies in the interface of closely packed transmembrane helices in the dimer structure of the activated receptor (S. O. Smith, unpublished results). It has been proposed that the transmembrane glycine in the *neu/erbB-2* receptor is part of a more general five-residue motif that occurs in receptor tyrosine kinase transmembrane helices where the second residue in the motif is either glycine or alanine (Sternberg and Gullick, 1989, 1990). In a similar manner, small residues line the faces of transmembrane helices in viral coat proteins (Deber et al., 1993) and in the α- and β-chains of the major histocompatibility complex (Cosson and Bonifacino, 1992). In single-pass membrane proteins, glycine residues are thought to facilitate helix packing either by forming favorable van der Waals surfaces for hydrophobic packing or by allowing closer dipolar interactions of the polar backbone atoms. These studies raise the question of whether similar interactions occur in polytopic membrane proteins.

The crystal structures of cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center of *Rhodobacter sphaeroides*, and the potassium channel of *Streptomyces lividans* clearly reveal that glycine residues

Received for publication 8 February 1999 and in final form 26 May 1999.
Address reprint requests to Steven O. Smith, Ph.D., Dept. of Biochemistry and Cell Biology, 312 Life Sciences Bldg., Nicholls Rd., SUNY Stony Brook, Stony Brook, NY 11794-5215. Tel.: 516-632-1210; Fax: 516-632-8575; E-mail: steven.o.smith@sunysb.edu.

0006-3495/99/09/1609/10 $2.00

are tolerated in transmembrane helices. Together these four membrane proteins have 48 transmembrane helices containing a total of 104 glycine residues. The occurrence of glycine in the transmembrane helices of the proteins studied (~8.7%) is almost identical to the overall occurrence of glycine in soluble proteins (~9%) (Nakashima et al., 1986). In a comparison of the amino acid composition of 24 membrane proteins, both transport and non-transport proteins, it has been shown that the transmembrane helices actually have a higher proportion of glycine residues relative to their aqueous domains (Deber et al., 1986).

In this paper we characterize the nature and distribution of the amino acids in the interfaces between transmembrane helices and in the region of helix crossings. A detailed analysis of helix packing is presented for cytochrome c oxidase. Fig. 1 shows the position of the 28 transmembrane helices in the crystal structure of cytochrome c oxidase viewed from above the membrane plane. The transmembrane glycine residues are highlighted in red and scattered throughout the sequence. Helix contact plots reveal that glycine is favored between transmembrane helices, particularly in the region of helix crossings. In the four proteins, a good correlation is found between the molecular volume of an amino acid and its occurrence in a helix interface. Glycine and alanine have the largest positive preferences for an interior orientation and are found with high frequency at the point of closest packing between the transmembrane helices. Together, these observations are relevant for understanding how membrane proteins fold in hydrophobic environments and suggest that glycine interactions have a structural role in helix packing similar to that observed in single-pass membrane proteins.

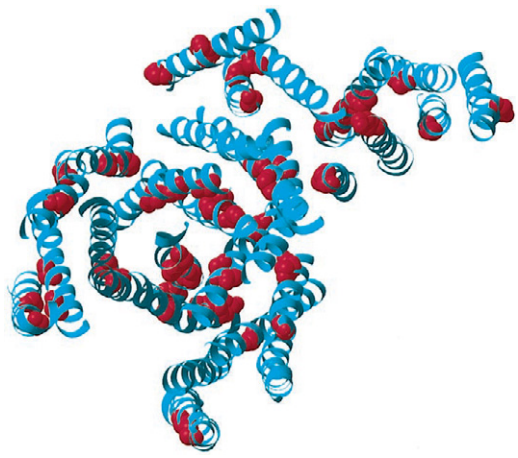

FIGURE 1 Crystal structure of cytochrome c oxidase. The protein complex has 13 different subunits containing a total of 28 transmembrane helices with relatively few chromophores or bound metal ions. As a result, the tertiary fold of the protein is dominated by helix-to-helix packing interactions. The transmembrane helices are shown with glycine residues highlighted in red.

METHODS

The coordinates for cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center of *R. sphaeroides*, and the potassium channel of *S. lividans* were taken from the Brookhaven Protein Data Bank (PDB access numbers 1occ, 2brd, 1aij, and 1bl8, respectively). The central hydrophobic portion of the 48 transmembrane helices in these proteins is bracketed by polar, charged, and aromatic residues. For the calculations below, we assigned the hydrophobic boundaries based on the position of basic and acidic residues. This yields an average hydrophobic helix length of 22 residues. The N-terminus of helix 7 in subunit III of cytochrome c oxidase was not defined by an Arg, Lys, His, Glu, or Asp residue. The analysis described below is not significantly influenced by the exact assignment of the transmembrane region of the helices.

Helix packing was evaluated using a modified version of the program Euler (Peersen, 1994) which calculates backbone to backbone distances between transmembrane helices (i.e., between all backbone amide N, Cα, and carbonyl C atoms). The modified program identifies the secondary structure of a protein and then calculates the interatomic distances between all backbone atoms of one helix and all backbone atoms of another helix. The program reduces the vast amount of structural data to an understandable table of results. In our analysis two helices are considered to be interacting if the minimum backbone-to-backbone distance is between 3 and 8 Å and there are at least 100 distances of $<$8 Å between heteroatoms. In the region of contact for closely packed helices, there are an average of ~10–12 backbone-backbone distances for each interfacial atom. As an example, Fig. 2 presents a contact plot for two interacting helices in cytochrome c oxidase. Helix 1 of subunit VIIb is packed against helix 1 of subunit IV. For each backbone atom on helix 1(VIIb), distances are calculated to all backbone atoms on helix 1(IV). From this matrix of data, the shortest interhelical backbone-to-backbone distance is plotted for each residue on helix 1(VIIb). The plot has an oscillation of 3.6 residues, since the distances are calculated as one moves along the peptide sequence. The

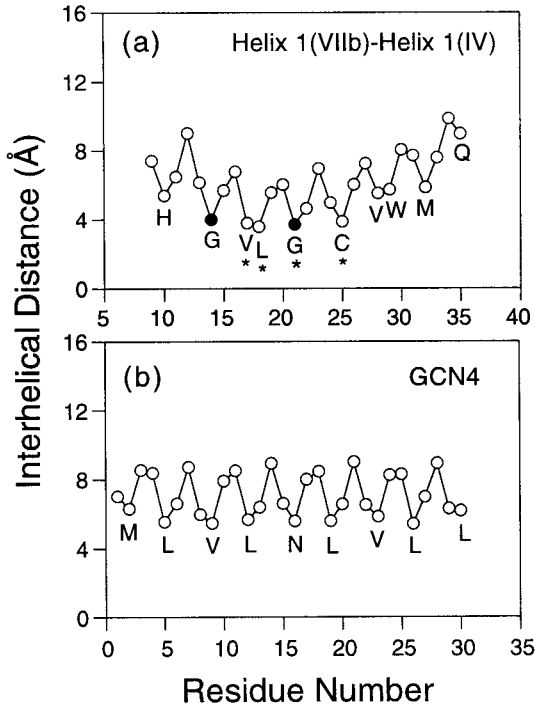

FIGURE 2 Helix packing curves of backbone-to-backbone distances. (*a*) Helix 1 of subunit VIIb is packed against helix 1 of subunit IV in cytochrome c oxidase. Glycine residues in helix 1 (VIIb) are indicated by filled circles. Asterisks indicate those residues on helix 1 (VIIb) that pack against glycine residues on helix 1 (IV). (*b*) Helices in the leucine zipper motif of GCN4.

helices pack in a parallel fashion with a left-handed crossing angle of 24° (Table 1). These helices, like many in polytopic membrane proteins, do not pack as coiled coils. A helix-packing curve for the leucine zipper motif of GCN4, the prototypical left-handed coiled coil, is shown in Fig. 2 *b*. The interhelical distance is roughly constant as one moves along the sequence. This is in contrast to the left-handed crossing geometry for helices 1(VIIb) and 1(IV), where the distance is a minimum roughly at the center of the sequence and then diverges toward the ends.

Crossing angles between closely packed helices were calculated using the program define_s (Richards and Kundrot, 1988). The crossing angle for any given helix pair was calculated in the region of the minimum in a helix-helix contact plot. Both left- and right-handed crossing angles are observed in the four membrane protein structures. The backbone dihedral angles psi (Ψ) and phi (Φ) were measured using the graphics program InsightII (Molecular Simulations, San Diego, CA).

RESULTS

Occurrence of glycine in membrane-spanning helices

Glycine residues are more or less evenly distributed throughout the structure of the four membrane proteins studied in this paper. In cytochrome c oxidase, 25 of the 28 transmembrane helices contain glycine (56 total). Remarkably, proline residues are also relatively common; 14 of the 28 transmembrane helices contain proline (21 total). Only a single helix, helix 1 of subunit II, contains neither a glycine nor a proline. Bacteriorhodopsin contains seven transmembrane helices that surround a central retinal prosthetic group. Four of the helices contain a total of 13 glycine residues. Glycine, leucine, and methionine are the most frequently occurring amino acids at the crossing points of the transmembrane helices. The photosynthetic reaction center contains 11 transmembrane helices and several bound chromophores. All of the helices contain at least one glycine (30 total). Alanine and glycine are the predominant amino acids at the helix crossing points. The potassium channel is the simplest of the helical transmembrane proteins studied here. The channel is formed from identical subunits, each containing two transmembrane helices containing two and three glycines, respectively. Alanine and valine have the highest occurrence in the helix interfaces.

Glycine residues that are present in the membrane-spanning helices of the proteins studied do not disrupt the secondary structure. In cytochrome c oxidase 13 of the helices contain both proline and glycine, yet retain their helical secondary structure. Fig. 3 presents the Φ and Ψ dihedral angles for glycine and proline residues in cytochrome c oxidase and shows that they fall in the standard α-helical region of a Ramachandran plot. In cytochrome c oxidase, bends are not seen in helices that contain a single proline residue in the middle of the transmembrane helix. However, bends are observed in helices that contain more than one proline residue or a combination of proline and glycine residues spaced four residues apart. Such examples are seen in helices 6(I), 2(II), and 1(VIIc) where bends of 19°, 20°, and 36° are observed, respectively. In helices 6(I) and 2(II), there are two proline residues present in the middle of the transmembrane helix, and in helix 1(VIIc) the bend is attributed to the presence of one proline and two glycine residues in the transmembrane helix.

TABLE 1 Interacting helices in cytochrome c oxidase

Left-Handed				Right-Handed			
Parallel		Antiparallel		Parallel		Antiparallel	
Helices	Angle	Helices	Angle	Helices	Angle	Helices	Angle
2(I)–4(I)	11°	2(I)–3(I)*	11°	5(I)–7(I)*	−8°	10(I)–11(I)*	−11°
9(I)–11(I)	14°	3(III)–6(III)	16°	3(III)–7(III)	−9°	6(I)–7(I)*	−18°
8(I)–1(II)	15°	4(III)–1(VIa)	17°	1(VIIc)–1(VIII)*	−15°	6(I)–3(III)*	−19°
1(I)–3(I)	17°	4(III)–7(III)	18°	3(I)–1(III)	−28°	1(II)–1(VIc)	−20°
6(I)–8(I)	19°	11(I)–12(I)*	18°	5(I)–3(III)	−46°	9(I)–1(II)	−23°
1(IV)–1(VIIb)*	24°	4(III)–5(III)	20°	11(I)–1(IV)	−49°	2(I)–1(VIIc)*	−24°
2(I)–10(I)	33°	5(III)–6(III)*	20°	1(I)–1(VIIc)*	−50°	8(I)–2(II)*	−25°
1(III)–1(VIIa)*	39°	1(I)–2(I)*	21°			12(I)–1(VIIc)	−25°
2(I)–6(I)	40°	1(II)–2(II)	21°			12(I)–1(VIII)*	−40°
6(I)–10(I)*	49°	1(III)–2(III)*	21°			4(I)–3(III)*	−41°
		6(III)–7(III)*	21°			12(I)–1(IV)	−55°
		3(I)–4(I)	22°				
		5(I)–6(I)*	22°				
		8(I)–9(I)	24°				
		7(I)–8(I)	25°				
		9(I)–10(I)*	27°				
		1(I)–12(I)	32°				
		4(I)–5(I)	32°				
		2(III)–1(VIIa)	38°				
		1(I)–10(I)	41°				
		6(I)–9(I)	45°				

*Helix pairs having a glycine residue at the closest point of interaction.

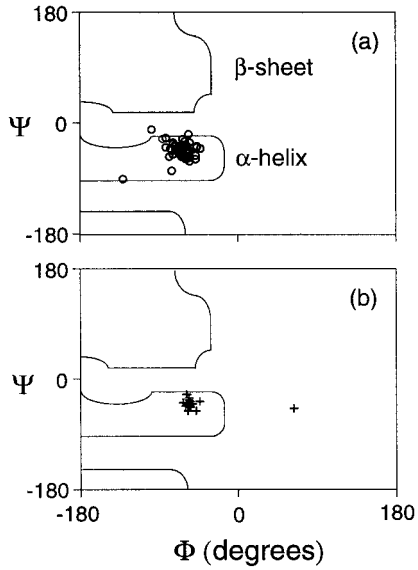

FIGURE 3 Dihedral Ψ and Φ angles for glycine (*a*) and proline (*b*) in the 28 transmembrane helices of cytochrome c oxidase.

Analysis of helix-helix interactions using helix-packing curves

Helix-packing curves are used to analyze the helix-to-helix interactions in cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center, and the bacterial potassium channel. Helix-to-helix packing interactions dominate the transmembrane structure of these proteins. Table 1 summarizes the orientations and angles between the most closely packed helices of the multi-subunit protein complex of cytochrome c oxidase. These helix pairs have right- and left-handed crossing angles with both parallel and antiparallel orientations. Left-handed crossing angles are favored over right-handed angles, and antiparallel orientations over parallel orientations (Bowie, 1997a).

The packing interactions among the 28 helices of cytochrome c oxidase range from closely packed coiled coils to helices that have only a few van der Waals contacts. In the 50 helix pairs of cytochrome c oxidase, 20 contain at least one glycine within the helix interface (Table 1). One way to visualize packing interactions between transmembrane helices is to plot the minimum distance between the helix backbones, as shown in Fig. 2. Residues that are positioned in the helix-helix interface appear at the local minima in the packing curves. The global minimum represents the crossing point of two helices. As described in the Methods section this analysis was done for all four of the proteins studied. The transmembrane helices can pack with either left- or right-handed crossing angles. The helix packing curves in Figs. 4 and 5 provide examples where glycine is located at both local and global minima.

Fig. 4 presents the pairwise interaction of four of the 28 helices in the cytochrome c oxidase complex. Helices 1, 2, and 5 of subunit I and helix 1 of subunit VIIc all have transmembrane glycine residues. Gly-32 in helix 1(VIIc) is

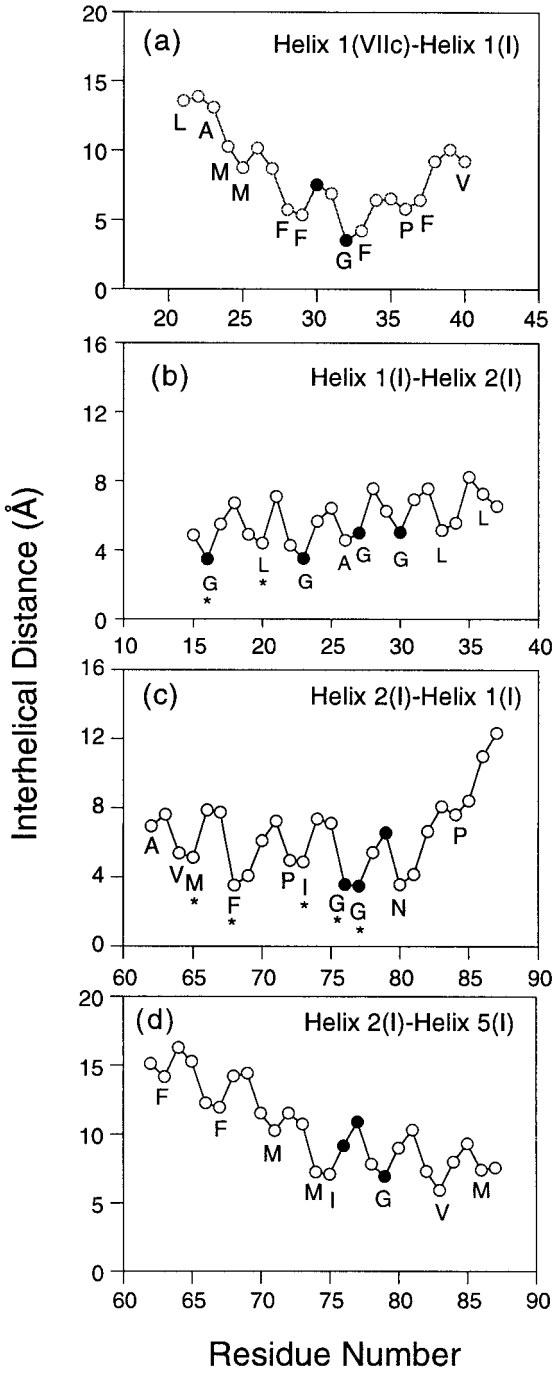

FIGURE 4 Helix-packing curves for cytochrome c oxidase. (*a*) Helices 1(VIIc)-1(I), (*b*) helices 1(I)-2(I), (*c*) helices 2(I)-1(I), and (*d*) helices 2(I)-5(I). Glycine residues are indicated by filled circles. Asterisks indicate those residues that pack against glycine residues on the opposing helix.

at the point of closest contact with helix 1(I) (Fig. 4 *a*). This glycine is flanked by phenylalanine residues creating a "molecular notch" on the helix surface into which Val-29 on helix 1(I) packs. These two helices diverge from the point of closest approach and represent a clear example of a helix-helix cross.

Helix 1(I) of cytochrome c oxidase provides an example of a transmembrane helix having two distinct packing sur-

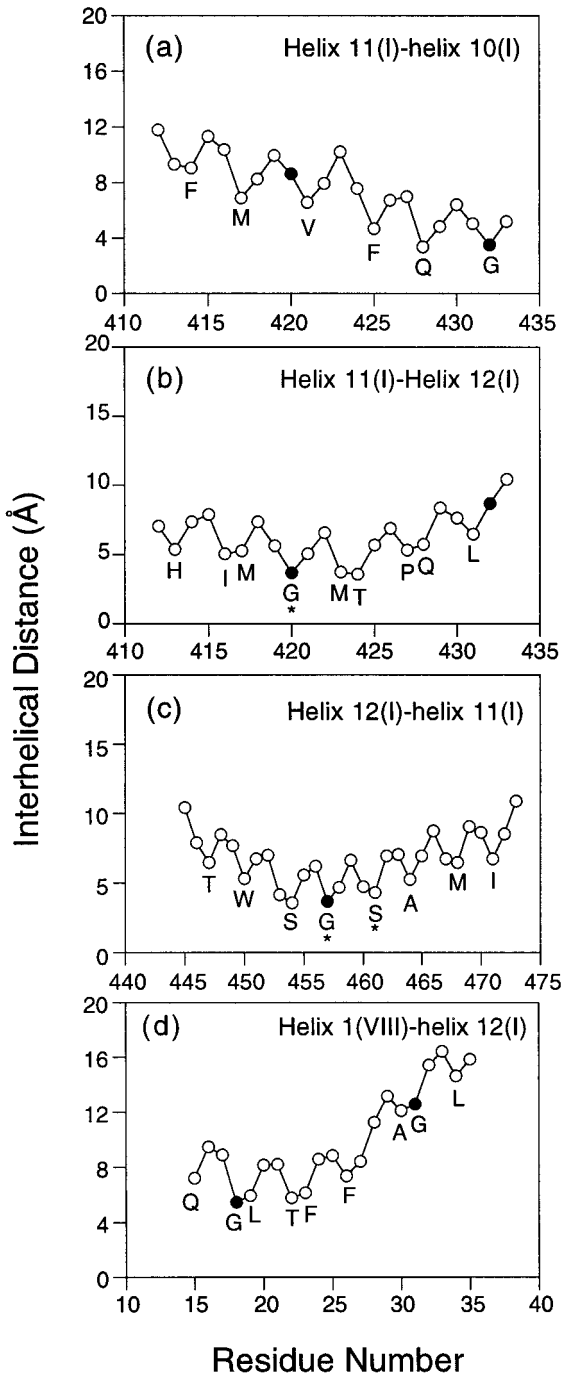

FIGURE 5 Helix-packing curves for cytochrome c oxidase. (*a*) Helices 11(I)-10(I), (*b*) helices 11(I)-12(I), (*c*) helices 12(I)-11(I), and (*d*) helices 1(VIII)-12(I). Glycine residues are indicated by filled circles. Asterisks indicate those residues that pack against glycine residues on the opposing helix.

faces. The surface of helix1(I) containing Val-29 is lined with bulky residues that pack against glycine residues on the opposing helix (helix 1(VIIc)). The second packing surface of helix 1(I) is lined by a series of glycine residues (Fig. 4 *b*). The first glycine residue in the sequence, Gly-16, packs against another glycine on helix 2(I). This is marked by an asterisk in Fig. 4 *b*. The series of glycine residues lying on one face of a transmembrane helix is more typical of single-pass membrane proteins than polytopic membrane proteins (see below).

Glycine-glycine packing provides for the closest approach of the helix backbones. Several glycine-glycine packing interactions are observed among the helix pairs of the four proteins we studied. A good example of glycine-glycine interactions is seen in the packing curves of helix 2(I) and helix 1(I) of cytochrome c oxidase (Fig. 4 *c*). In the helix 2(I) packing curve in Fig. 4 *c*, a third glycine residue (Gly-79) is oriented away from the interface, but is seen to lie in the helix interface with helix 5(I) in Fig. 4 *d*. The picture that emerges from these four interacting helices is one where the glycine residues serve as notches for orienting multiple helices in the folded complex.

To emphasize this concept of helix packing, Fig. 5 presents the pairwise interactions of four additional helices in the cytochrome c oxidase complex, i.e., helices 10 (I), 11(I), 12(I), and 1(VIII). Each helix pair shown has a glycine residue at or near the global minimum. In Fig. 5, *b* and *c*, glycine-glycine packing is observed in the interaction of helices 11(I) and 12(I). These two helices in turn pack against glycines in two more helices, 10(I) and 1(VIII), shown in Fig. 5, *a* and *d*, forming a tightly packed bundle of helices. In these two packing curves, the glycine residues at the global minima interact with the backbone atoms of aromatic residues tyrosine and phenylalanine.

Glycine occurs in regions of helix-to-helix contact

The helix-packing curves suggest that glycine residues facilitate helix packing in membrane proteins. The quantitative analysis to test this suggestion is presented in this section. Fig. 6 *a* shows the distribution of amino acids in the transmembrane helices of cytochrome c oxidase. This distribution generally correlates with hydrophobicity (Wallin et al., 1997). The most abundant amino acids are hydrophobic: leucine, valine, isoleucine, alanine, and phenylalanine. Polar residues, such as serine and threonine, are often found in the transmembrane region of membrane proteins, and have a relatively high occurrence in cytochrome c oxidase. The preference for glycine in the transmembrane region is close to that of serine and threonine, but is significantly greater than that of proline, tyrosine, and histidine.

Fig. 6 *b* presents the occurrence of amino acids at the global minimum when the packing curves for all pairwise interactions are analyzed among the 28 helices in the cytochrome c oxidase monomer (Table 1). The most striking feature of this plot is the high occurrence of glycine. This observation parallels the observation of glycine residues in the dimer interface of membrane proteins having only a single transmembrane helix (Lemmon and Engelman, 1994). The occurrence of glycine residues at the global minima is in part due to the lack of a bulky side chain.

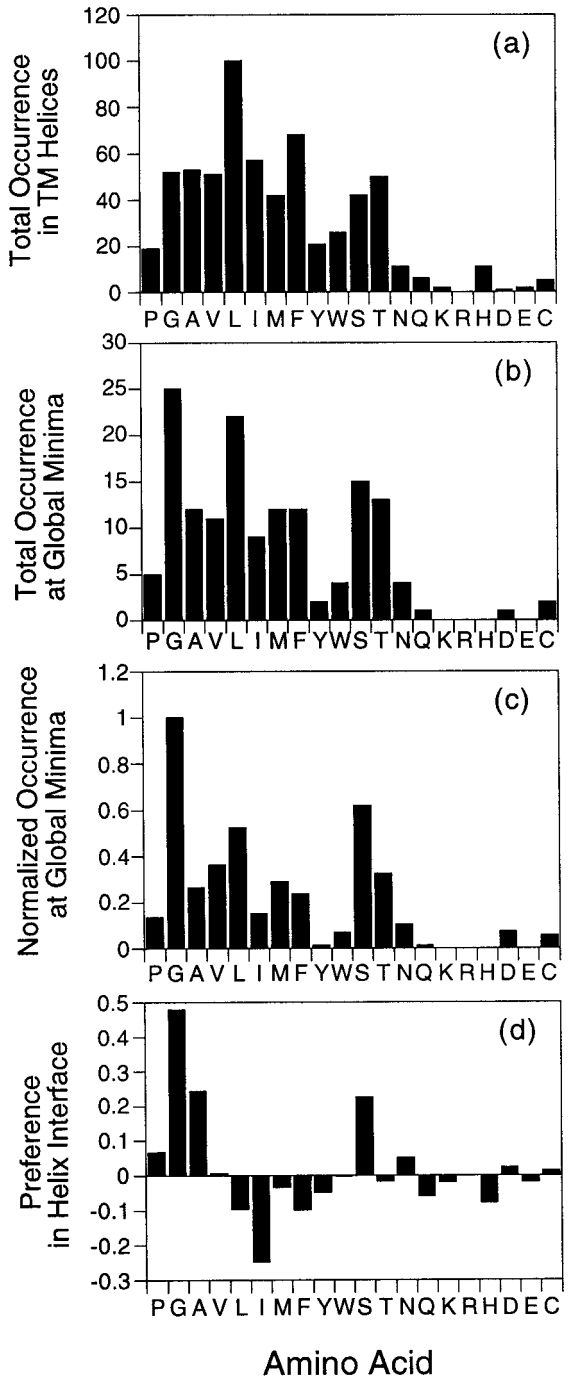

FIGURE 6 (*a*) Distribution of amino acid residues in the transmembrane helices of cytochrome c oxidase. (*b*) Total occurrence of amino acid residues at the point of closest helix-to-helix contact. The closest interhelical distance was calculated as described in the Methods section and in Fig. 2. Residues are counted as being at the closest helix-to-helix contact if they are at the absolute minimum of the helix contact plot or they are within 0.5 Å of the minimum. The inclusion of residues within 0.5 Å of the minimum provides for uncertainty in the 2.8 Å crystal structure while excluding adjacent residues in the sequence as also being in the interface. (*c*) Normalized occurrence of amino acid residues at the point of closest approach of transmembrane helices relative to the total number of amino acids of a given type in the transmembrane region and to the total number of residues at the global minima. (*d*) Preference for packing in a helix-helix interface.

However, this occurrence becomes more striking when one considers that 20 of the 50 helix pairs in cytochrome c oxidase have at least one glycine residue at the closest contact point (Table 1) and that 25 of the 52 transmembrane glycines are located at global minima. The large number of glycine residues at global minima is not due simply to the large number of glycine residues in the transmembrane helices. This is more likely the case for common membrane-spanning residues such as leucine. Glycine residues are mostly observed at the transmembrane helix interfaces, and only in two instances occur in a potential C-cap position (data not shown). This is in contrast to what is seen in soluble proteins where the C-cap position is overwhelmingly dominated by glycine residues, which terminate ~34% of the helices (Richardson and Richardson, 1988; Aurora et al., 1994). Fig. 6 *c* presents the number of residues at the global minimum in the helix-packing curves normalized to the total number of amino acids of a given type in the transmembrane region *and* to the total number of residues at the global minima. The predominance of glycine is striking.

Fig. 6 *d* presents a histogram that reflects the preference for an amino acid residue to pack in a helix-helix interface of cytochrome c oxidase. For this plot, the residues were defined as those that occur at *local* minima and had a backbone-backbone separation of 6 Å or less. This cutoff was based on the helix separation in GCN4, a well-packed left-handed coiled coil of helices (Fig. 2 *b*). The plot was generated by subtracting the normalized values of each amino acid residue in the transmembrane domain from the normalized values of the occurrence of each amino acid at the 6 Å interface. Positive values reflect a preference to lie in an interfacial position, while negative values reflect a preference to lie out of an interface (e.g., to be oriented toward the lipid-protein interface). Glycine, alanine, and serine have the largest positive preferences for an interior orientation.

Fig. 7, *a–c* combine the data obtained from the packing curves of cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center of *R. sphaeroides*, and the potassium channel of *S. lividans*. Fig. 7 *a* plots the occurrence of amino acids at the global minimum when the packing interactions of all pairwise interactions of these four membrane proteins are taken into account. The most striking feature of this plot is the high occurrence of alanine, glycine, and leucine residues. As in the analysis of cytochrome c oxidase alone, it is revealing to normalize these values to the total number of each residue type. This highlights the predominance of glycine (Fig. 7 *b*) as was the case in Fig. 6 *c*. The preference of an amino acid residue to pack in a helix-helix interface is shown in Fig. 7 *c*. This plot was generated in the same way as Fig. 6 *d* by subtraction of the normalized values of each amino acid from the normalized values of the occurrence of each amino acid at the 6 Å interface. Positive values reflect the preference of amino acids to lie in an interfacial position, whereas negative values reflect the preference of amino acid residues to be oriented toward the lipid-protein interface. Again, glycine

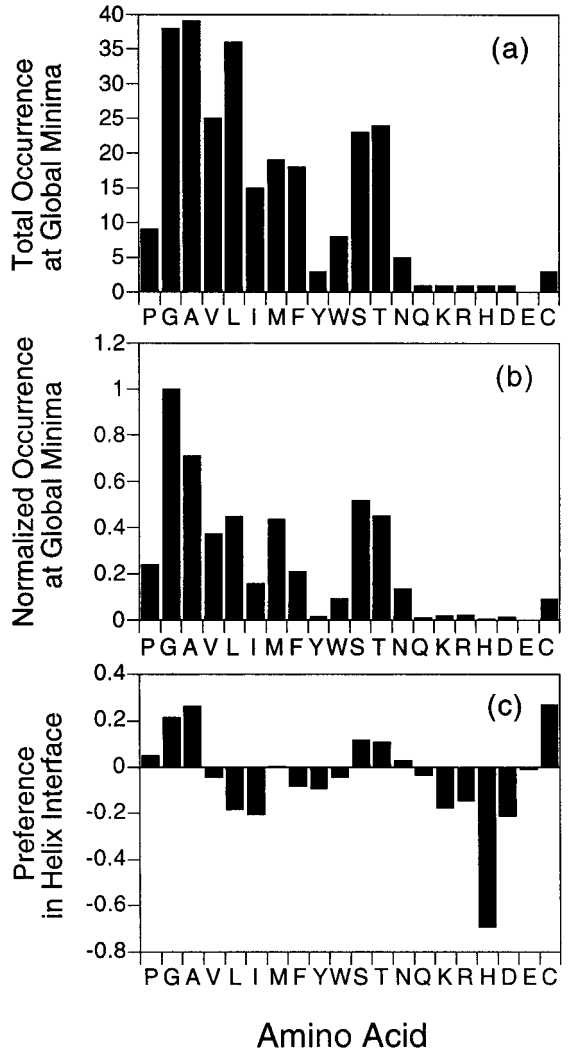

FIGURE 7 Combined analysis of amino acid residues in the transmembrane helices of cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center complex, and the bacterial potassium channel. (*a*) Total occurrence of amino acid residues at the point of closest helix-to-helix contact. The closest interhelical distance was calculated as described in the Methods section and in Fig. 2. (*b*) Normalized occurrence of amino acid residues at the point of closest approach of transmembrane helices relative to the total number of amino acids of a given type in the transmembrane region and to the total number of residues at the global minima. (*c*) Preference for packing in a helix-helix interface.

and alanine residues have a preference for an interior orientation, as was seen in the analysis of cytochrome oxidase alone.

In a statistical analysis of single-spanning membrane proteins, Hunt (1993) noted that glycine residues often occur on the same face of an α-helix with an average spacing of 3.6 residues per turn. This is in contrast to what is seen in polytopic membrane proteins. The difference in the alignment of glycine residues in single-spanning and polytopic membrane proteins is most likely attributed to the structural interactions of the transmembrane helices. Single-spanning membrane proteins most often participate in structural interactions involving only one face of the helix (i.e., to form a dimer). For example, in glycophorin A two glycine residues are separated by four residues and are an essential element of the dimerization motif (Lemmon et al., 1994). In contrast, polytopic membrane proteins participate in structural interactions involving more than one face of the helix, as shown in Figs. 4 and 5 above.

Occurrence of interfacial amino acids as a function of molecular volume

Fig. 8 shows an inverse correlation between the molecular volume of an amino acid and its occurrence in an α-helical interface of a membrane protein. Interestingly, the correlation holds for residues at the helix crossing points (global minima), as well as the helix-helix interfaces (local minima) (data not shown). Glycine, alanine, and serine residues, which have the smallest volumes, have a tendency to be oriented toward the helix-helix interface of the four membrane proteins studied (Fig. 7 *c*). Serine residues, with their polar β-hydroxyl group, have the potential to hydrogen-bond across the helix interface. Serine can also contribute to polar pockets within the hydrophobic protein interior. In cytochrome c oxidase, serine residues are components of the hydrophilic channels that have been proposed in subunit I for transporting protons through the protein (Tsukihara et al., 1996).

Glycine packing interactions at helix contact points

Glycine residues mediate helix packing through van der Waals interactions. There are two factors that contribute to favorable glycine interactions in interfaces, both of which result from the lack of a side chain. The first factor is simply

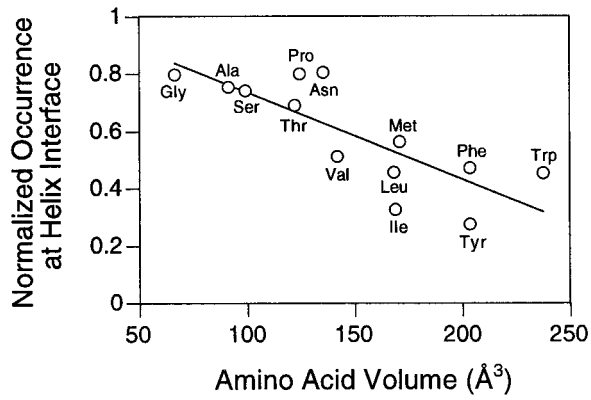

FIGURE 8 Normalized occurrence in a helix interface as a function of residue volume. The number of residues of each amino acid type was determined from the helix-packing curves of the helix pairs in cytochrome c oxidase, bacteriorhodopsin, the photosynthetic reaction center complex, and the bacterial potassium channel. Those residues were included that were at local minima and had a backbone-backbone distance of 6 Å or less. Those residues were excluded whose overall total occurrence in the transmembrane helices was $<1\%$ of the total number of amino acids in helix interfaces. The amino acid volumes were taken from Chothia (1975).

related to the complementarity between residue side chains. Residues with bulky side chains appear to pack well against glycine residues, and glycine themselves appear to pack well against glycines. Glycine has the highest average occluded surface of all residues in cytochrome c oxidase (S. C. Shekar, T. Shieh, P. J. Fleming, and S. O. Smith, submitted for publication) indicating packing voids are generally not formed at the position of glycine in the folded protein structure. In Fig. 2 *a*, the interaction between helix 1 on subunit VIIb and helix 1 on subunit IV of cytochrome c oxidase provides a typical example of both packing cases, namely bulky hydrophobic residues packing against glycines and glycines packing against glycines. In this example, Val-17 and Leu-18 on helix 1(VIIb) pack against two glycines on the helix 1(IV), while Gly-21 on helix 1(VIIb) also packs against a glycine residue on helix 1(IV). Such glycine-glycine packing is not uncommon. Two examples were illustrated above in Figs. 4 and 5. Fig. 9 *A* shows glycine interactions at the crossing point of helices 11(I) and 12(I).

The second factor that contributes to the favorable glycine interactions is that the absence of a side chain effectively exposes the polar peptide backbone of the helix. The backbone carbonyl and methylene groups bear substantial partial charges, and complementary charge interactions in a low dielectric membrane environment can represent a sizable driving force for helix association. In fact, the partial charges on the backbone carbonyl and NH groups themselves represent the driving force for the formation of helical secondary structure. The partial charge on methylene protons has led to the observation in soluble proteins of C-H · · · O hydrogen bonds (Derewenda et al., 1994, 1995). This is illustrated in Fig. 9 *B* for the interaction between Gly-420 on helix 11(I) and Gly-457 on helix 12(1). The helices are slightly offset such that the carbonyl group of Gly-420 is opposite the methylene group of Gly-457.

Fig. 4, *b* and *c* show the packing curves for helix 1 and helix 2 of subunit I of cytochrome c oxidase. The molecular packing of this interaction is seen in Fig. 9 *C*. The protons of Gly-16 on helix 1(I) bear partial positive charge and are packed close to the backbone carbonyl oxygens of helix 2(I) that have a significant partial negative charge. Additionally, the protons of Gly-76 on helix 2(I) are packed against the aromatic ring of Tyr-19 on helix 1(I). This also represents a dipole-dipole interaction if one considers the negative partial charge associated with the aromatic ring current. This type of glycine-aromatic interaction is also seen in the structure of bacteriorhodopsin (Grigorieff et al., 1996).

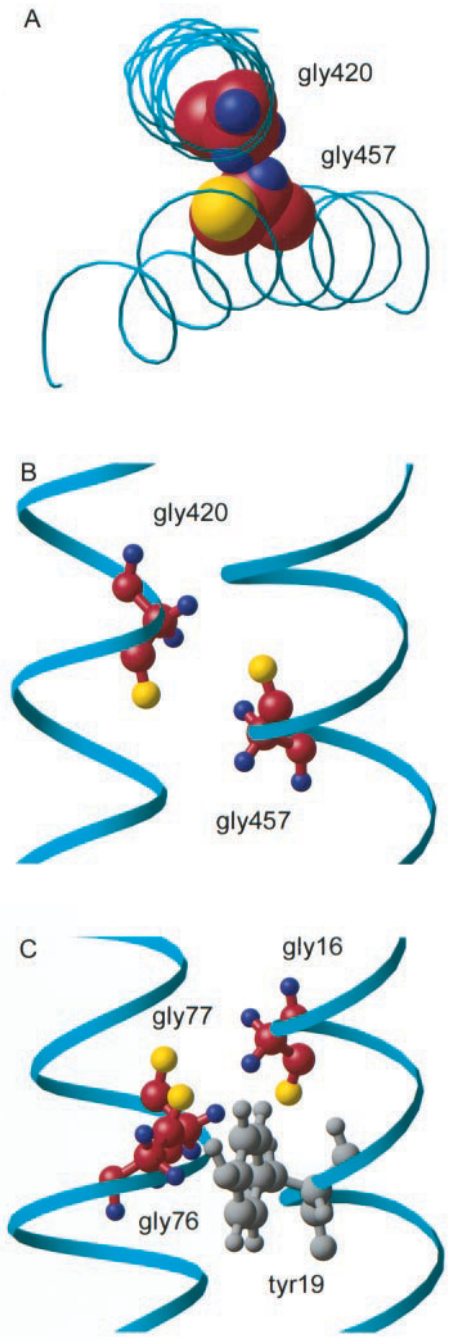

FIGURE 9 Molecular packing of glycine residues in cytochrome c oxidase. (*A*) Molecular packing of Gly-420 on helix 11 with Gly-457 of helix 12 of subunit I. (*B*) Molecular packing of Gly-420 on helix 11 with Gly-457 of helix 12 of subunit I. (*C*) Molecular packing of Gly-16 on helix 1 with Gly-76 and Gly-77 on helix 2 of subunit I.

DISCUSSION

The structural basis for how proteins fold in membranes may turn out to be simpler than for proteins in solution. Most membrane proteins span the bilayer with long α-helical stretches of amino acids. The formation of helical structure is due to the energetics of hydrogen-bonding of the backbone carbonyl and amide groups in the hydrophobic membrane interior. The protein folding problem in membranes is reduced to understanding how the transmembrane helices pack. Packing is largely driven by enthalpic rather than entropic interactions. The hydrophobic effect, which plays a dominant role in the folding of soluble proteins, does not contribute to helix association in membrane proteins once the helices are inserted into the lipid bilayer.

The packing of transmembrane helices has best been described for dimers of single-pass membrane proteins. Glycophorin A spans red blood cell membranes with a single α-helix and was shown by Marchesi and co-workers to dimerize through transmembrane interactions (Bormann et al., 1989). The sequence motif responsible for the dimerization of glycophorin A is LIxxGVxxGVxxT. Membrane protein chimera containing this motif retain the dimerization specificity of glycophorin A (Lemmon et al., 1994). Replacement of either glycine in the motif with other hydrophobic residues, even as small as alanine, disrupts dimer formation (Lemmon et al., 1992b). The glycines in the dimerization motif are oriented toward the helix interface. Similarly, the transmembrane oligomerization segments of the class II major histocompatibility complex (MHC) α and β chains (Cosson and Bonifacino, 1992), the signal anchor sequence of mitochondrial Mas70p (Millar and Shore, 1993), and the *neu/erbB-2* receptor tyrosine kinase (Sternberg and Gullick, 1989; 1990; Burke et al., 1997) contain small residues that are central to the sequence-specific dimerization of these proteins.

Helix packing in the membrane proteins studied here tends to support the idea that glycine residues facilitate helix interactions. The lack of a side chain provides a good packing surface and exposes the polar backbone. Glycine residues are the most occluded residues in membrane proteins, indicating that they generally do not produce packing voids in protein interiors (S. C. Shekar, T. Shieh, P. J. Fleming, and S. O. Smith, submitted for publication). In fact, of the 10 most highly occluded residues in cytochrome c oxidase (with residue packing values ranging from 0.604 to 0.671), eight occur in the region of closest approach between helices, and six are glycine. All together, glycines account for 12% of the residues in the helix interfaces of the four membrane proteins studied. In comparison, glycine represents only ~4% of the residues at helix interfaces in soluble proteins (Chothia et al., 1981). The high occurrence of glycine residues at helix crossing points is also likely to contribute to the broad distribution of crossing angles in the transmembrane helices of membrane proteins. In cytochrome c oxidase this range is from 8° to 55° (Table 1). These values differ from those observed by Chothia and co-workers who used the "ridges and grooves" model to predict the crossing angles of soluble proteins (Chothia et al., 1977, 1981). The predicted and observed crossing angles in soluble proteins are ~23° and −52° for left-handed and right-handed coiled coils, respectively. Recently, Bowie (1997a, b) studied the helix-packing distribution for transmembrane helices and concluded that the crossing angle deviation seen in transmembrane helices relative to the predicted values for soluble proteins is due to the size of the residues at the helix interface and the packing flexibility of transmembrane helices. In an analysis of helix-helix packing in soluble proteins, Walther et al. (1996) found a correlation among the size of the residues in the helix interface, the interaxial separation, and the preferred packing cell. The distribution of interhelical distances in the most closely packed cell peaked at 7.5 Å and was dominated by glycine and alanine in the interface (Walther et al., 1996).

The detailed analysis of the preference of the different amino acids to line helix interfaces extends the work of Bowie (1997a, b), Wallin et al. (1997), and many others. Most studies are generally consistent with the conclusion that small residues, both polar and hydrophobic, are favored in helix-helix interfaces. One notable exception is the conclusion of Wallin et al. (1997) that alanine is strongly preferred in exposed orientations rather than in helix interfaces, as indicated by Figs. 6 *d* and 7 *c*. Both the distance matrix approach described above and an occluded surface approach (S. C. Shekar, T. Shieh, P. J. Fleming, and S. O. Smith, submitted for publication) indicate that alanine, on-average, lies in well-packed interior positions. The observation that the small polar residues (Ser, Thr, and Asn) are generally favored in helix interfaces is consistent with these residues having interior orientations where they can participate in interhelical hydrogen-bonding interactions. A detailed analysis of the packing interactions of Ser, Thr, and Asn of the type described here for glycine will provide answers to the roles these residues have in membrane protein structure and function.

Glycine also appears to be important in the structure and function of polytopic membrane proteins whose structures have not been solved to high resolution. For example, lac permease is a polytopic membrane protein that catalyzes the coupled translocation of β-galactosides and protons. The protein has 12 transmembrane helices. Kaback and co-workers have systematically replaced 34 of the 36 glycine residues in lac permease with cysteine (Frillingos et al., 1997). Of these, three replacements (Gly-64, Gly-115, and Gly-147) in the transmembrane helices resulted in complete loss of activity. Further hydrophobic replacements at these sites revealed that the increased residue volume at the position of the glycine was responsible for inactivation. A second example involves Ca^{2+}-ATPase. There are three glycine residues (Gly-310, Gly-770, and Gly-801) in the putative transmembrane helices of this protein that play a major role in the active transport of Ca^{2+}. Anderson and co-workers have used site-specific mutagenesis to replace these glycines with either alanine or valine residues (Andersen et al., 1992). Their results suggest that these glycine residues are involved in Ca^{2+} binding and/or the enzyme conformational changes responsible for ion translocation. Finally, glycine residues may also facilitate van der Waals packing in the binding of ligands within membrane proteins. The vertebrate photoreceptor rhodopsin has a number of key transmembrane glycine residues. One of the most intensely studied is Gly-121, which forms part of the retinal binding pocket and is thought to form a cavity for packing of the C19 methyl group of the retinal (Han et al., 1996, 1997). Substitution of larger hydrophobic residues for Gly-121 leads to steric clashes and dark activation of the receptor (Han et al., 1996). This observation suggests that glycine may play a role in ligand binding similar to that in helix-helix packing.

In summary, glycine appears to have a structural role in membrane proteins that is distinct from that in soluble proteins. Glycine does not disrupt the secondary structure of helical transmembrane segments, but rather functions as a molecular notch to facilitate helix packing. By taking into account the location of glycine residues in membrane proteins one may better understand the nature of the interactions that guide helix association in hydrophobic membrane environments and stabilize membrane protein structure.

We are grateful to the crystallographers who deposit their coordinates in the Protein Data Bank.

This work was supported by National Institutes of Health Grants GM 46732 and GM 41412.

REFERENCES

Andersen, J. P., B. Vilsen, and D. H. MacLennan. 1992. Functional consequences of alterations to Gly-310, Gly-770, and Gly-801 located in the transmembrane domain of the Ca(2+)-ATPase of sarcoplasmic reticulum. *J. Biol. Chem.* 267:2767–2774.

Aurora, R., R. Srinivasan, and G. D. Rose. 1994. Rules for α-helix termination by glycine. *Science*. 264:1126–1130.

Bormann, B. J., W. J. Knowles, and V. T. Marchesi. 1989. Synthetic peptides mimic the assembly of transmembrane glycoproteins. *J. Biol. Chem.* 264:4033–4037.

Bowie, J. U. 1997a. Helix packing in membrane proteins. *J. Mol. Biol.* 272:780–789.

Bowie, J. U. 1997b. Helix packing angle preferences. *Nat. Struct. Biol.* 4:915–917.

Burke, C. L., M. A. Lemmon, B. A. Coren, D. M. Engelman, and D. F. Stern. 1997. Dimerization of the p185neu transmembrane domain is necessary but not sufficient for transformation. *Oncogene*. 14:687–696.

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. USA*. 74:4130–4134.

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

Cosson, P., and J. S. Bonifacino. 1992. Role of transmembrane domain interactions in the assembly of class II MHC molecules. *Science*. 258: 659–662.

Deber, C. M., C. J. Brandl, R. B. Deber, L. C. Hsu, and X. K. Young. 1986. Amino acid composition of the membrane and aqueous domains of integral membrane proteins. *Arch. Biochem. Biophys.* 251:68–76.

Deber, C. M., A. R. Khan, Z. Li, C. Joensson, M. Glibowicka, and J. Wang. 1993. Val→Ala mutations selectively alter helix-helix packing in the transmembrane segment of phage M13 coat protein. *Proc. Natl. Acad. Sci. USA*. 90:11648–11652.

Derewenda, Z. S., U. Derewenda, and P. M. Kobos. 1994. (His)C^{ϵ}-H $\cdots$ O═C hydrogen bond in the active sites of serine hydrolases. *J. Mol. Biol.* 241:83–93.

Derewenda, Z. S., L. Lee, and U. Derewenda. 1995. The occurrence of C-H $\cdots$ O hydrogen bonds in proteins. *J. Mol. Biol.* 252:248–262.

Frillingos, S., M. L. Ujwal, J. Sun, and H. R. Kaback. 1997. The role of helix VIII in the lactose permease of *Escherichia coli*. I. Cys-scanning mutagenesis. *Protein Sci.* 6:431–437.

Grigorieff, N., T. A. Ceska, K. H. Downing, J. M. Baldwin, and R. Henderson. 1996. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* 259:393–421.

Han, M., M. Groesbeek, T. P. Sakmar, and S. O. Smith. 1997. The C9 methyl group of retinal interacts with glycine-121 in rhodopsin. *Proc. Natl. Acad. Sci. USA*. 94:13442–13447.

Han, M., S. W. Lin, S. O. Smith, and T. P. Sakmar. 1996. The effects of amino acid replacements of glycine 121 on transmembrane helix 3 of rhodopsin. *J. Biol. Chem.* 271:32330–32336.

Hunt, J. F. 1993. Biophysical studies of integral membrane protein folding. Ph.D. thesis. Yale University, New Haven, CT.

Landolt-Marticorena, C., K. A. Williams, C. M. Deber, and R. A. Reithmeier. 1993. Non-random distribution of amino acids in the transmembrane segments of human type I single span membrane proteins. *J. Mol. Biol.* 229:602–608.

Lemmon, M. A., and D. M. Engelman. 1994. Specificity and promiscuity in membrane helix interactions. *Quart. Rev. Biophys.* 27:157–218.

Lemmon, M. A., J. M. Flanagan, J. F. Hunt, B. D. Adair, B. J. Bormann, C. E. Dempsey, and D. M. Engelman. 1992a. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J. Biol. Chem.* 267:7683–7689.

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

Lemmon, M. A., H. R. Treutlein, P. D. Adams, A. T. Brünger, and D. M. Engelman. 1994. A dimerization motif for transmembrane alpha-helices. *Nat. Struct. Biol.* 1:157–163.

Li, S. C., and C. M. Deber. 1992. Glycine and β-branched residues support and modulate peptide helicity in membrane environments. *FEBS Lett.* 311:217–220.

MacKenzie, K. R., J. H. Prestegard, and D. M. Engelman. 1997. A transmembrane helix dimer: structure and implications. *Science*. 276: 131–133.

Millar, D. G., and G. C. Shore. 1993. The signal anchor sequence of mitochondrial Mas70p contains an oligomerization domain. *J. Biol. Chem.* 268:18403–18406.

Nakashima, H., K. Nishikawa, and T. Ooi. 1986. The folding type of a protein is relevant to the amino acid composition. *J. Biochem. (Tokyo).* 99:153–162.

O'Neil, K. T., and W. F. DeGrado. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*. 250:646–651.

Peersen, O. B. 1994. Rotational resonance NMR determination of $^{13}C \cdots {}^{13}C$ distances in crystalline and membrane bound proteins. Ph.D. Thesis. Yale University, New Haven, CT. 139–144.

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.

Richardson, J. S., and D. C. Richardson. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science*. 240:1648–1652.

Smith, S. O., and B. J. Bormann. 1995. Determination of helix-helix interactions in membranes by rotational resonance NMR. *Proc. Natl. Acad. Sci. USA*. 92:488–491.

Sternberg, M. J., and W. J. Gullick. 1989. Neu receptor dimerization. *Nature*. 339:587.

Sternberg, M. J., and W. J. Gullick. 1990. A sequence motif in the transmembrane region of growth factor receptors with tyrosine kinase activity mediates dimerization. *Protein Eng.* 3:245–248.

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.

Wallin, E., T. Tsukihara, S. Yoshikawa, G. Von Heijne, and A. Elofsson. 1997. Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. *Protein Sci.* 6:808–815.

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.