

# The C $\alpha$ —H $\cdots$ O hydrogen bond: A determinant of stability and specificity in transmembrane helix interactions

Alessandro Senes\*, Iban Ubarretxena-Belandia\*, and Donald M. Engelman†

Department of Molecular Biophysics and Biochemistry, Yale University, P.O. Box 208114, New Haven, CT 06520

Contributed by Donald M. Engelman, June 5, 2001

The C $\alpha$ —H $\cdots$ O hydrogen bond has been given little attention as a determinant of transmembrane helix association. Stimulated by recent calculations suggesting that such bonds can be much stronger than has been supposed, we have analyzed 11 known membrane protein structures and found that apparent carbon  $\alpha$  hydrogen bonds cluster frequently at glycine-, serine-, and threonine-rich packing interfaces between transmembrane helices. Parallel right-handed helix–helix interactions appear to favor C $\alpha$ —H $\cdots$ O bond formation. In particular, C $\alpha$ —H $\cdots$ O interactions are frequent between helices having the structural motif of the glycoporphin A dimer and the GxxxG pair. We suggest that C $\alpha$ —H $\cdots$ O hydrogen bonds are important determinants of stability and, depending on packing, specificity in membrane protein folding.

The hydrogen bond is a key element in the interplay between stability and specificity in protein folding. The desolvation penalty associated with burial of polar side chains in an aqueous environment is not always fully recovered by hydrogen bond formation, so hydrogen bonds provide a small or even unfavorable net energy contribution to folding. However, the strength and directionality of hydrogen bonds make them an important factor in discriminating between correctly folded and misfolded states. Hence, polar interactions tend to contribute more to specificity than to stability in soluble proteins (1–3). Conversely, in the apolar environment of biological membranes donor and acceptor groups cannot be satisfied by the solvent, and hydrogen bonds strongly stabilize the helical conformation of membrane spanning domains (4) and can stabilize tertiary interactions as well (5–9). We are interested in the role of hydrogen bonds in the association of transmembrane helices, a stage that is pivotal in the folding of membrane proteins (4). Recently, the DeGrado group and our laboratory showed that the substitution of a single polar amino acid residue into model transmembrane helices induces homooligomerization (10–13); the association driven by hydrogen bonding can be strong and independent of packing details. Thus, in the apolar environment, the strength of hydrogen bonds can stabilize the association of transmembrane helices, although a lack of a need for sequence specificity could create a danger of inducing promiscuous association (10, 13).

Weaker hydrogen bonds, such as those between carbon and oxygen atoms (C—H $\cdots$ O), have received little attention in the membrane protein field, and their occurrence in membrane proteins has never been surveyed. The C $\alpha$  is an activated carbon donor because it is bound to the electron-withdrawing amide N—H and C=O groups, and, in soluble proteins, hydrogen bonds between main-chain C $\alpha$ —H groups and backbone or side-chain oxygen atoms are often observed (14–17). Despite its abundance, the structural contribution of the C $\alpha$ —H $\cdots$ O hydrogen bond has been unclear and its interaction energy has been believed to be small. Recently, by using *ab initio* calculations, Vargas *et al.* (18) and Scheiner *et al.* (19) estimated the energy of the C $\alpha$ —H $\cdots$ O hydrogen bond to be as much as 2.5–3.0 kcal/mol *in vacuo*, placing the strength of the C $\alpha$ —H $\cdots$ O hydrogen bond at approximately one-half the energy of a “conven-

**Table 1. Structural database of nonhomologous helical membrane proteins**

Protein	Protein Data		Helix–helix interactions*†	C $\alpha$ —H $\cdots$ O contacts‡
	Base	Resolution, Å		
Bacteriorhodopsin	1c3w	1.55	9	13 (6)
Calcium ATPase	1eul	2.60	15	11 (4)
Cytochrome c oxidase	1occ	2.80	48	53 (12)
Fumarate reductase	1qla	2.20	6	5 (2)
Glycerol facilitator	1fx8	2.20	9	19 (11)
Glycophorin A	1afo	NMR	1	12 (6)
Light-harvesting complex II	1lgh	2.40	1	0
Mechanosensitive channel	1msl	3.50	3	0
Photosynthetic reaction center	1prc	2.30	13	19 (4)
Potassium channel	1bl8	3.20	7	7 (3)
Rhodopsin	1f88	2.80	13	6 (3)
Total			125	145 (51)

\*Helical segments as identified by the DSSP program (21).

†Unique interactions in homo-oligomers.

‡Number of contacts with  $d_H < 3.5$  Å and  $\zeta > 120^\circ$  (number of contacts with  $d < 2.7$  Å).

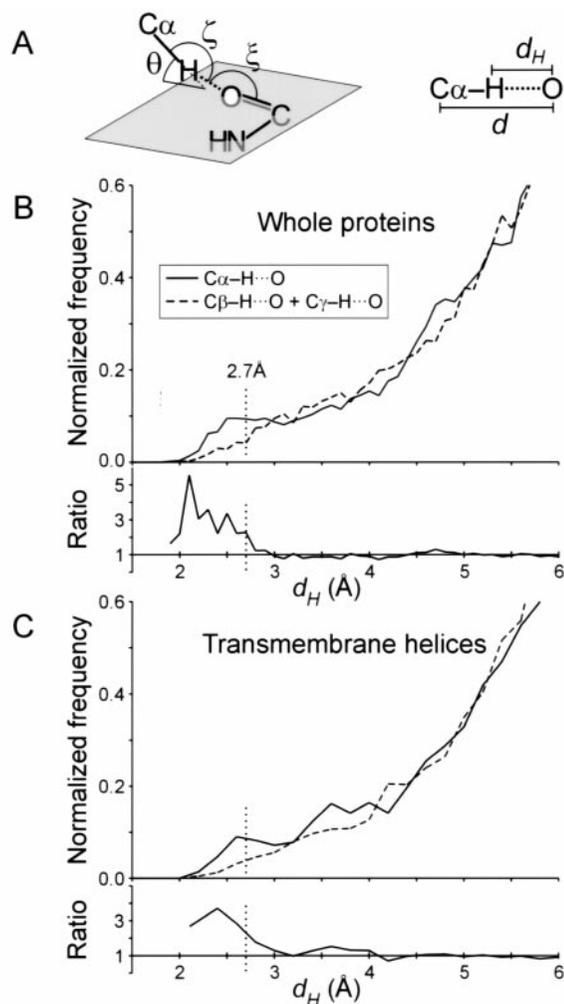
ditional” N—H $\cdots$ O hydrogen bond. Given the importance of electrostatic interactions in the apolar interior of a membrane lipid bilayer, this energy may be significant for providing stability in transmembrane helix–helix association, especially when several C $\alpha$ —H $\cdots$ O interactions are coordinated at a single interface. To examine this hypothesis, we have analyzed 125 helix–helix interfaces in a database of 11 nonhomologous membrane protein structures (Table 1) in search of C $\alpha$ —H $\cdots$ O contacts that display the geometric hallmarks of hydrogen bonds. The analysis is limited by the uncertainty in the positioning of atoms in crystal structures solved at 2- to 3-Å resolution, although even at these resolutions the position of backbone atoms and helix axes should be quite accurate. We find recurring patterns when C $\alpha$ —H $\cdots$ O contacts cluster at interfaces between helices that display a short interaxial distance, have a preference for parallel right-handed packing, and are rich in glycine, serine, and threonine residues. Our results suggest that C $\alpha$ —H $\cdots$ O hydrogen bonds are overlooked factors that can determine stability and, given their dependence on the packing details, also specificity in the interaction of transmembrane helices.

Abbreviations: GpA, glycoporphin A; GlpF, glycerol facilitator.

\*A.S. and I.U.-B. contributed equally to this work.

†To whom reprint requests should be addressed. E-mail: don@chimera.csb.yale.edu.

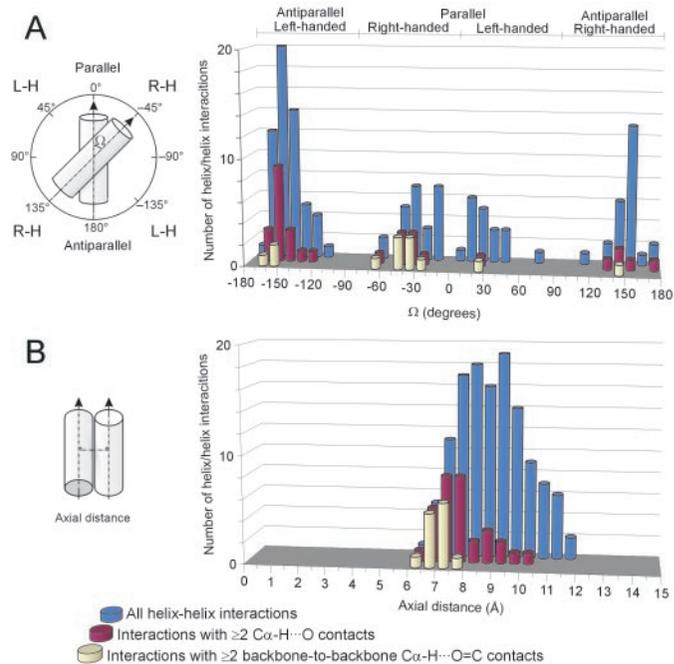
The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.



**Fig. 1.** (A) Definition of the geometrical parameters of the C $\alpha$ -H...O hydrogen bond. Nomenclature according to Derewenda *et al.* (14). The ideal values (14, 18) are as follows: H—O distance,  $d_H \leq 2.7$  Å; C $\alpha$ —O distance,  $d \leq 3.8$  Å; C $\alpha$ —H—O angle,  $\zeta = 180^\circ$ ; H—O—C angle,  $\xi = 120^\circ$ ; elevation angle,  $\theta = 0^\circ$  (angle between the C $\alpha$ —H vector and the amide plane). (B) Distribution of hydrogen to acceptor distances ( $d_H$ ) for all C $\alpha$ —H donor groups (solid line), compared with that of C $\beta$ —H + C $\gamma$ —H groups (dashed line), in the 11 membrane proteins structures. The ratio between the two curves is also shown (C $\alpha$ /C $\beta\gamma$ ). The control set is similarly sized and is composed by the C $\beta$ —H groups of all Ile, Leu, Val, Met, Phe, and Tyr residues and the C $\gamma$ —H of Leu and C $\gamma$ 1—H of Ile. Contacts below 2.7 Å have an overall frequency of occurrence that is 3 times higher for C $\alpha$  donors than for the C $\beta\gamma$  control set. (C) Analogous distribution as in B, but limited to the subset of residues in helical transmembrane segments.

## Methods

Our database of 11 nonhomologous helical membrane proteins is listed in Table 1. Hydrogen atoms were added to the crystallographic protein database coordinate files with the program REDUCE (20). Identification of the helical segments was performed with the program DSSP (21). For homo-oligomers, only a single representative of the duplicated helix-helix interaction was considered. The structural analysis of C $\alpha$ -H...O contacts (distances  $d_H$  and  $d$  and angles  $\zeta$ ,  $\xi$ , and  $\theta$ , nomenclature according to Derewenda *et al.* (14), defined in Fig. 1A) and the calculations of interhelical distances and angles were performed with a custom-made PERL program (available from the authors on request). Interhelical angles and axial distances were calculated at the point of minimal axial distance by using the local



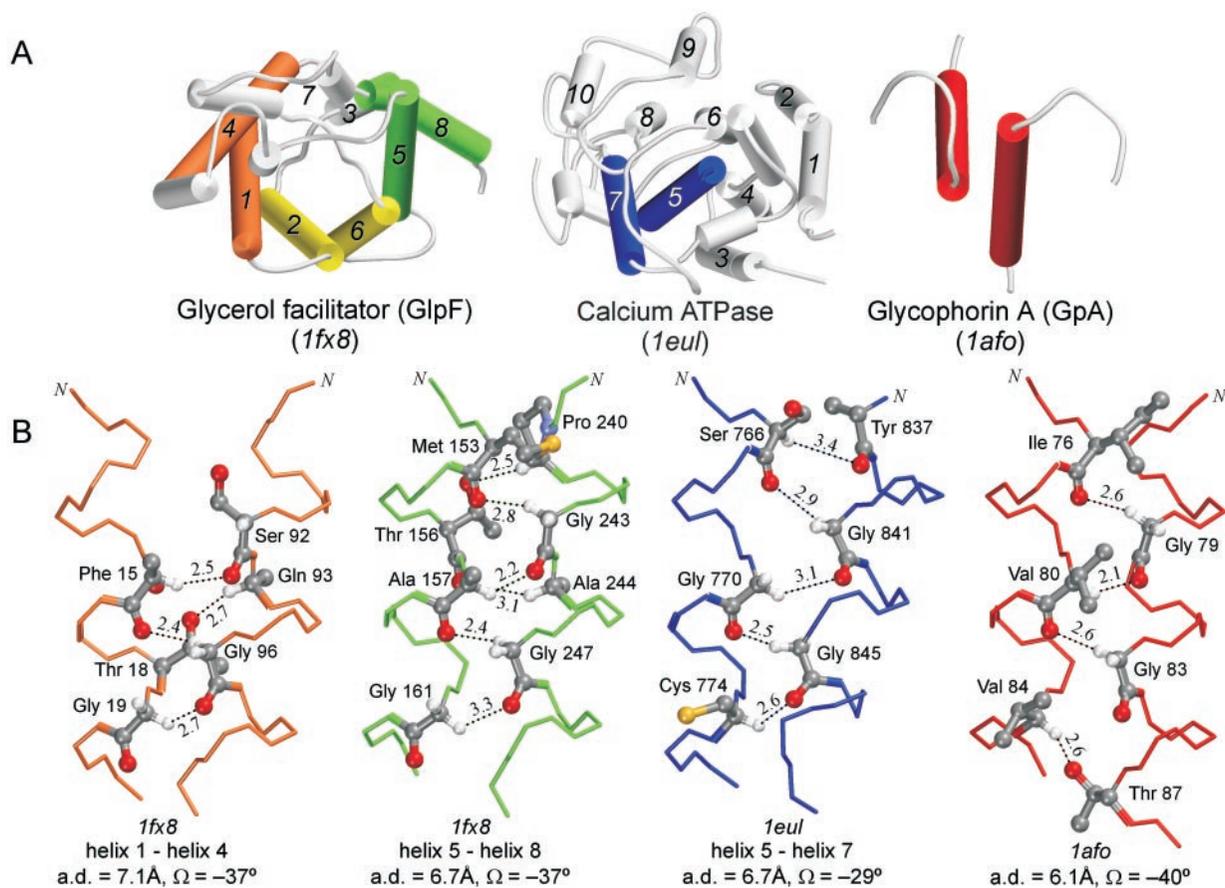
**Fig. 2.** (A) Distribution of interhelical packing angles ( $\Omega$ ). Interhelical packing angles were calculated as the angle between the local helical axes at the point of minimal axial distance. L-H, left-handed; R-H, right-handed; blue, all helix-helix interactions; red, helix-helix interactions with at least two C $\alpha$ -H...O contacts to backbone or side-chain oxygen atoms; yellow, helix-helix interactions with at least two backbone-to-backbone C $\alpha$ -H...O=C contacts. (B) Distribution of interhelical axial distances among the same three categories as for A.

helical axis. The local helical axes were calculated by using the coordinates of four contiguous C $\alpha$  atoms according to Sugeta and Miyazawa (22) with a PERL subroutine adapted from the FORTRAN program HELANAL (23).

## Results and Discussion

**Distribution of Interhelical C $\alpha$ -H...O Hydrogen Bonds in Transmembrane Helices.** The optimal geometry for a strong C $\alpha$ -H...O hydrogen bond is described in Fig. 1A. The H...O distance ( $d_H$ ) should be smaller than 2.7 Å (the sum of the van der Waals radii), and all three atoms should be aligned (C—H—O angle  $\zeta = 180^\circ$ ) (14, 17, 18). We calculated the  $d_H$  distribution of all C $\alpha$  hydrogen donors to backbone and side chain acceptors in our database of membrane proteins and compared it to that obtained with a similarly sized set of aliphatic C $\beta$  and C $\gamma$  hydrogen atoms (Fig. 1B). The ratio of the two distributions shows that C $\alpha$  hydrogen atoms form contacts below van der Waals separation more frequently (overall, 3 times more frequently below 2.7 Å). This is consistent with the fact that the C $\alpha$  is an activated carbon donor with a higher tendency to form C—H...O hydrogen bonds. Despite the lower sample size, a similar trend is observed when the analogous distribution is calculated limited to the subset of donors and acceptor groups of transmembrane helical segments (Fig. 1C).

We studied the geometry of interaction of all potential C $\alpha$ -H...O hydrogen bonds between transmembrane helical segments. Hydrogen bond interactions persist at a longer range than van der Waals separation and tolerate significant angular distortion (17). Operatively, we selected  $d_H < 3.5$  Å and  $\zeta > 120^\circ$  (or  $\zeta > 90^\circ$  when  $d_H < 3.0$  Å) as a comprehensive limit for recording C $\alpha$ -H...O contacts as potential hydrogen bonds. We found 145 interhelical C $\alpha$ -H...O contacts, 51 of which have  $d_H$



**Fig. 3.** Parallel right-handed helix-helix interactions with extended networks of  $C_\alpha$ -H $\cdots$ O contacts (GpA-like motifs). (A) Schematic representation of the structure of the glycerol facilitator (GlpF, Protein Database ID 1fx8), the calcium ATPase (1eul), and GpA (model 19 in 1afo). The page is parallel to the plane of the membrane. The color coding corresponds to the interactions shown in B. The interaction between helices 2 and 6 in GlpF (yellow) is antiparallel and is shown in Fig. 4A. Some extramembranous regions of the calcium ATPase were removed for clarity. (B) Apparent networks of  $C_\alpha$ -H $\cdots$ O hydrogen bonds at the interface of four right-handed helix-helix interactions. Some side-chain atoms have been removed for clarity. In the GpA homodimer (red) each interaction occurs symmetrically on both sides but only one set of interactions is shown for clarity. Apparent hydrogen bonds are denoted with dots ( $\cdots$ ) and the distance ( $d_H$ ) in Å is indicated. The interhelical axial distance (a.d.) and packing angle ( $\Omega$ ) of the helix-helix interactions are also indicated.

$< 2.7$  Å in 125 helix-helix interactions between 103 helices (Table 1). About one-fourth of all helix-helix interactions contain at least two instances of interhelical  $C_\alpha$ -H $\cdots$ O contacts to either backbone or side-chain oxygen acceptors. One-tenth contain at least two backbone-to-backbone  $C_\alpha$ -H $\cdots$ O=C contacts. The distribution of these subsets of helix-helix interactions as a function of packing angle ( $\Omega$ ) is displayed in Fig. 24. In the total set of interactions (blue) we found a strong bias for left-handed packing, as in an earlier analysis by Bowie (24). The bias is mainly because of the antiparallel component (57 left- vs. 25 right-handed). When one considers only those helix-helix interactions with a minimum of two  $C_\alpha$ -H $\cdots$ O contacts to backbone or side-chain acceptors (red), some preference for right-handed parallel and left-handed antiparallel interactions appears. When a further requirement of at least two backbone-to-backbone  $C_\alpha$ -H $\cdots$ O=C contacts is introduced (yellow), a dramatic selection for right-handed parallel interactions is seen ( $\chi^2$  test,  $P < 0.005$ ). Notably, in the  $-50^\circ$  to  $-20^\circ$  range of packing angles,  $\approx 50\%$  of all helix-helix interactions persist after the selection. Thus, parallel right-handed packing favors interhelical backbone-to-backbone contacts and  $C_\alpha$ -H $\cdots$ O formation.

The analysis of the occurrence of  $C_\alpha$ -H $\cdots$ O contacts as a function of the interhelical axial distance is shown in Fig. 2B. All interhelical distances in the database are between 6 and 12 Å

with an average of 8.9 Å. The average drops to 7.8 Å when at least two  $C_\alpha$ -H $\cdots$ O contacts are present and it is further reduced to 7.0 Å when at least two backbone-to-backbone contacts are present (with no instance above 7.6 Å). Notably, below 7.0 Å all helix-helix interactions contain potential  $C_\alpha$ -H $\cdots$ O=C backbone-to-backbone bonds.

**Parallel Right-Handed Helix-Helix Interactions: Occurrence of Glycophorin A (GpA)-Like Motifs with Multiple  $C_\alpha$ -H $\cdots$ O Bonds.** Four cases of parallel right-handed helix-helix interactions with extended  $C_\alpha$ -H $\cdots$ O contacts are present in the database (Fig. 3). Two cases are found in the glycerol facilitator (GlpF) (25); a third example is observed in the calcium ATPase (26); and the fourth is found in the GpA dimer (27). The geometries of the  $C_\alpha$ -H $\cdots$ O contacts are shown in Table 2 together with the ideal distance and angle values.  $C_\alpha$ -H $\cdots$ O hydrogen bonds can tolerate some divergence from ideality (18), and steric factors imposed by helical packing may prevent ideal geometry. The four helix-helix interactions display very similar packing angles ( $-29^\circ$  to  $-40^\circ$ ). The interactions also have similar interfacial residues: pairs of small residues spaced at  $i, i + 4$  are present on all helices and, in particular, GxxxG motifs (28–30) are observed in all but one interface (a SxxxG motif is found in helix 4 of GlpF). As discussed later, the GxxxG motif is known to drive transmembrane helix association.

**Table 2. Geometry\* of C $\alpha$ —H $\cdots$ O contacts in GpA-like motifs**

Donor	Acceptor	$d_H$ , Å	$d$ , Å	$\zeta$ , deg	$\xi$ , deg	$\theta$ , deg
Ideal values						
C—H	O=C	$\leq 2.7$	$\leq 3.8$	180	120	0
Glycophorin A (1afo): model 19 <sup>†</sup> (average of 20 models <sup>‡</sup> )						
G79—2H $\alpha$	I76—O	2.6 (3.5)	3.6 (4.5)	156	119	37
V80—H $\alpha$	G79—O	2.1 (2.5)	2.9 (3.3)	129	102	71
G83—2H $\alpha$	V80—O	2.6 (2.8)	3.7 (3.9)	170	112	40
V84—H $\alpha$	T87—O $\gamma$	2.6 (2.7)	3.5 (3.6)	140	133	—
Glycerol facilitator (1fx8): helix 1—helix 4						
F15—H $\alpha$	S92—O	2.5	3.5	149	118	59
G19—1H $\alpha$	G96—O	2.6	3.3	122	103	45
Q93—H $\alpha$	E14—O	3.2	4.3	156	114	84
Q93—H $\alpha$	T18—O $\gamma$	2.7	3.6	142	94	—
G96—2H $\alpha$	F15—O	2.4	3.4	145	123	45
Glycerol facilitator (1fx8): helix 5—helix 8						
P240—H $\alpha$	T156—O $\gamma$	2.5	3.4	147	89	—
G243—2H $\alpha$	M153—O	2.8	3.5	124	113	47
A244—H $\alpha$	T156—O	3.1	4.1	148	104	82
A157—H $\alpha$	G243—O	2.2	3.0	123	127	50
G247—2H $\alpha$	A157—O	2.4	3.3	134	107	57
G161—1H $\alpha$	G247—O	3.3	3.9	116	105	34
Calcium ATPase (1eul): helix 5—helix 7						
A762—H $\alpha$	T837—O $\eta$	3.3	4.0	127	67	33
S776—H $\alpha$	T837—O	3.4	4.2	130	93	6
G770—1H $\alpha$	G841—O	3.1	3.8	116	93	13
C774—H $\alpha$	G845—O	2.6	3.2	117	120	5
G841—1H $\alpha$	S766—O	2.9	3.4	111	137	6
G845—1H $\alpha$	G770—O	2.5	3.3	129	91	10

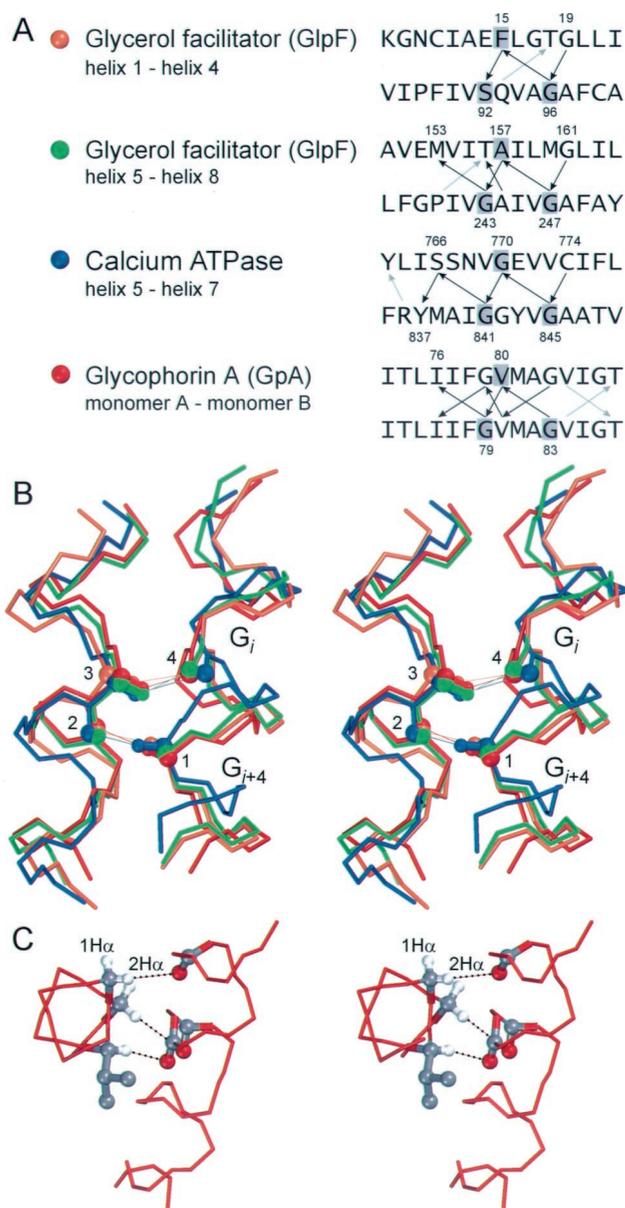
\*Nomenclature according to Derewenda *et al.* (14), defined in Fig. 1A.

<sup>†</sup>Average of the symmetric interactions on the two sides of the interface.

<sup>‡</sup>Average of all 40 symmetric interactions in the 20 NMR models.

Although the four right-handed structures have similar packing angles and interfacial residues, they, most strikingly, also share a pattern of C $\alpha$ —H $\cdots$ O connectivity. In Fig. 4A, the backbone-to-backbone C $\alpha$ —H $\cdots$ O=C contacts are represented by black arrows, and the backbone-to-side chain C $\alpha$ —H $\cdots$ O contacts by gray arrows. All backbone-to-backbone contacts occur with the same periodicity, between one residue and two residues spaced at  $i$ ,  $i + 4$  on the opposite helix. The  $i$ ,  $i + 4$  connectivity extends up to three helical turns. The GxxxG interaction motifs are central to this connectivity, as highlighted in the backbone superimposition of the four structures in Fig. 4B. The four connecting atoms used to align the structures are displayed in ball representation. On the right are shown the helices containing the GxxxG (SxxxG) motifs with the carbonyl oxygen of the Gly at  $i$  (4) and the C $\alpha$  of the Gly at  $i + 4$  (1); on the left, the C $\alpha$  (3) and the oxygen (2) of the residue that interacts with the two Gly residues. The interhelical packing angles of  $-29^\circ$  to  $-40^\circ$  appear to favor the  $i$ ,  $i + 4$  connectivity. The relative rotation that has to be applied to two opposing helices to align the vector joining the C $\alpha$  and the carbonyl oxygen of a residue on one helix (vector 2-3), with the vector joining a carbonyl oxygen with the C $\alpha$  at  $i - 4$  on the opposing helix (vector 1-4), is approximately  $-35^\circ$ .

Among the four structures, the GpA dimer is the only one with GxxxG pairs present on both interacting helix surfaces (Fig. 4A). GxxxG was identified by Russ and Engelman (29) as a major motif driving oligomerization in a screen of randomized interaction interfaces. Furthermore, small residues spaced at  $i$ ,  $i + 4$  have increased occurrence in transmembrane helices; in particular, GxxxG is the most biased pair in transmembrane helices, being strongly over-represented in both single-span and multi-span membrane proteins (28). Finally, several mutagenesis stud-



**Fig. 4.** (A) Connectivity of the networks of apparent C $\alpha$ —H $\cdots$ O hydrogen bonds in the four parallel right-handed GpA-like motifs. The arrows show the interactions in the donor-to-acceptor direction. Black arrows: backbone-to-backbone C $\alpha$ —H $\cdots$ O bonds; gray arrows: backbone-to-side chain bonds. Backbone-to-backbone C $\alpha$ —H $\cdots$ O=C contacts occur between an amino acid residue on one helix and two residues spaced at  $i$ ,  $i + 4$  on the opposite helix in all structures. The three residues involved in the alignment in B are shaded in the sequences. (B) Superimposition of the four structures aligned using the carbonyl oxygen (4) at  $i$  and the C $\alpha$  (1) at  $i + 4$  of the GxxxG (SxxxG) motifs on the helix on the right; and the C $\alpha$  (3) and carbonyl oxygen (2) of the residue that interacts with the two Gly residues with apparent C $\alpha$ —H $\cdots$ O bonds. The overall rms deviation of the superimposition calculated on the backbone atoms of 13 residues on each helix is 1.6 Å. (C) The 2H $\alpha$  of glycine residues is oriented roughly in the same direction of the H $\alpha$  of the residue at  $i + 1$  and  $i - 3$ . When a GxxxG motif is present a potential interaction interface arises. The example shows the interaction of Gly-79, Val-80, and Gly-83 of GpA that donate to carbonyl oxygen atoms spaced at  $i$ ,  $i + 3$  and  $i + 4$  on the opposite monomer.

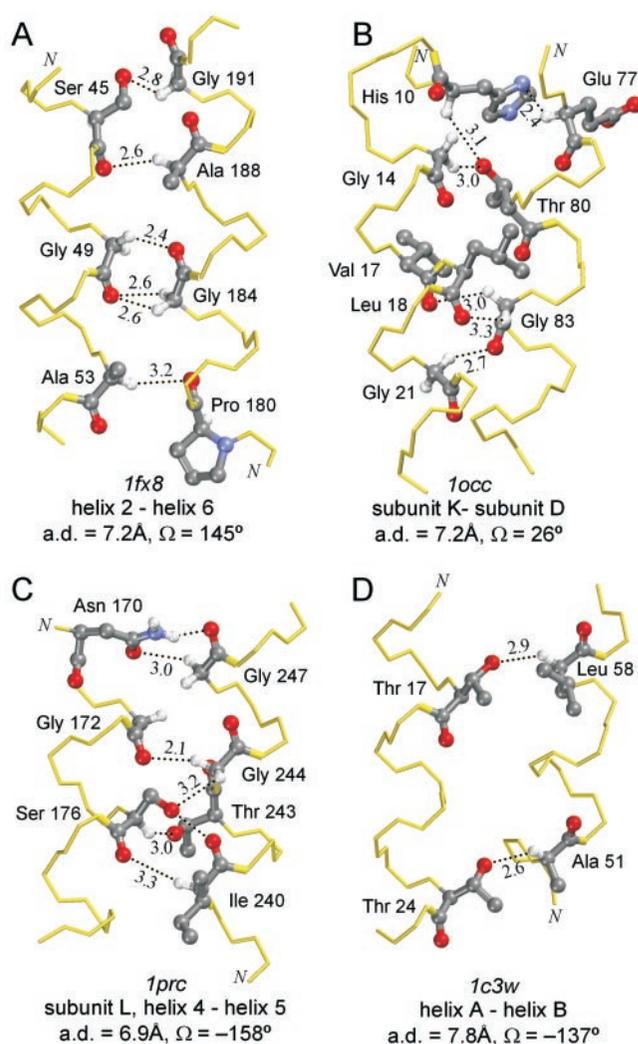
ies support the involvement of the GxxxG motif in transmembrane helix oligomerization and interaction (30–35). In addition to favoring interhelical backbone contacts because of the lack of a side chain, Gly residues also increase the opportunities for

$C\alpha-H\cdots O$  formation with two  $\alpha$ -hydrogen atoms that can act as donors (denoted as  $1H\alpha$  and  $2H\alpha$ ). In a helical conformation, the  $2H\alpha$  of Gly (the hydrogen that is stereochemically in the position of the side chain in other amino acids) points in the same direction as the  $H\alpha$  of the residues at  $i + 1$  and  $i - 3$ . Thus, in a  $GXXXG$  motif three  $H\alpha$  atoms (specifically, the  $2H\alpha$  atoms of each Gly residue, and the  $H\alpha$  of the residue X at  $i + 1$  with respect to the first Gly) are oriented roughly parallel and a potential interaction surface arises. This is illustrated in Fig. 4C, where the  $C\alpha-H$  of Gly-79, Val-80, and Gly-83 of GpA are donors to carbonyl oxygen atoms on the opposite side spaced at  $i, i + 3$  and  $i + 4$ . A similar pattern is also found in GlpF (donors Gly-243, Ala-244, and Gly-247). Hence, the present data suggest that the  $GXXXG$  motif drives transmembrane helix association in part by favoring  $C\alpha-H\cdots O=C$  interactions. We have found  $GXXXG$  pairs only in parallel right-handed “GpA-like” interaction motifs, although numerous interfacial Gly residues occur in all identified interhelical networks with backbone-to-backbone contacts. It will be interesting to see whether the expectation of finding  $GXXXG$  pairs frequently—and perhaps mainly—occurring in GpA-like motifs will be met once more structural data become available.

**Other Networks of Interhelical  $C\alpha-H\cdots O$  Bonds.** Fig. 5 shows four more examples of helix–helix interfaces with apparent  $C\alpha-H\cdots O$  bonds: three very extensive networks (A–C) containing several backbone-to-backbone  $C\alpha-H\cdots O$  contacts occur in an antiparallel right-handed interaction (A, GlpF), a parallel left-handed interaction (B, cytochrome *c* oxidase), and an antiparallel left-handed interaction (C, photosynthetic reaction center). The interaction of Fig. 5A is the third case in GlpF. As with the other two interactions, it is right-handed and characterized by small interfacial amino acid residues spaced at  $i, i + 4$  engaged in  $C\alpha-H\cdots O$  contacts ( $AxxxGxxxS$  on helix 2 and  $GxxxA$  on helix 6), but it is antiparallel. An example from bacteriorhodopsin (bR) is shown in Fig. 5D. The structure of bR has been solved at 1.55 Å and is currently the highest-resolution structure for a membrane protein. bR has interhelical distances in the range 7.8–10.6 Å, above the 7.6 Å that appears to be the approximate limit for backbone-to-backbone  $C\alpha-H\cdots O=C$  formation. However, backbone-to-side-chain  $C\alpha-H\cdots O$  bonds are observed.

It should be noted that the examples of Fig. 5 do not have the thematic packing angles and connectivity seen in the right-handed parallel interactions. Instead, the central element characterizing all observed cases with backbone-to-backbone  $C\alpha-H\cdots O=C$  networks is the presence of numerous interfacial Gly, Ser, and Thr residues.

**The Versatility of Glycine in Interhelical  $C\alpha-H\cdots O$  Bond Networks.** Gly residues are frequent in transmembrane helices, constituting 8% of the amino acid composition (28). Gly residues permit short interhelical separation, and it has previously been suggested that they might participate in  $C\alpha-H\cdots O$  bonds (36). In our analysis, 23% of all helical Gly residues appear to be involved as donors and 10% as acceptors in  $C\alpha-H\cdots O$  contacts. As previously discussed, Gly residues increase the opportunity for  $C\alpha-H\cdots O$  bond formation because the second  $H\alpha$  of glycine points roughly in the same direction as the  $H\alpha$  of the residues at  $i + 1$  and  $i - 3$  (Fig. 4C). In addition, both the  $H\alpha$  atoms of a Gly residue can simultaneously form  $C\alpha-H\cdots O$  bonds to either different acceptors (Gly-83 in Fig. 5B; Gly-244 in Fig. 5C) or the same acceptor oxygen (bidentate interactions, Gly-184 in Fig. 5A). Hence, the versatility of Gly residues appears to favor  $C\alpha-H\cdots O$  network formation.



**Fig. 5.** Additional helix–helix interactions with multiple  $C\alpha-H\cdots O$  hydrogen bonds at various interhelical packing angles. (A) Antiparallel right-handed interaction from the GlpF. (B) Parallel left-handed interaction from bovine heart cytochrome *c* oxidase. (C) Antiparallel left-handed interaction from the photosynthetic reaction center. (D) Antiparallel left-handed interaction with apparent backbone-to-side-chain  $C\alpha-H\cdots O$  bonds from bacteriorhodopsin.

**Serine and Threonine Allow  $C\alpha-H\cdots O\gamma$  Contacts at Longer Interhelical Distances.** Ser and Thr residues are frequently involved in interhelical  $C\alpha-H\cdots O$  bonds. They each constitute 5% of the amino acid composition of transmembrane helices (28), where they are well tolerated because the donor potential of their polar side chains can be satisfied by forming  $O-H\cdots O$  hydrogen bonds to the carbonyl at  $i - 4$  or  $i - 3$  on the same helix (37, 38). For this reason they have a weak tendency to form interhelical  $O-H\cdots O$  hydrogen bonds (11, 13). The  $O\gamma$ , however, is available as an acceptor and is displaced from the helix axis. In our database 24% of all Ser and 20% of all Thr residues appear to be involved in  $C\alpha-H\cdots O\gamma$  bonds. In particular, in GpA, Thr-87 allows the formation of an additional  $C\alpha-H\cdots O$  when the interhelical distance is too large for the  $C\alpha-H$  of Val-84 to reach the backbone oxygen on the opposite helix (Fig. 3B). Consistently, the isosteric mutation of Thr-87 to Val results in partial destabilization of the GpA dimer (32). Moreover, the  $GXXXGXXXT$  motif was found by Russ and Engelman to be among the strongly associating helices in their selection of randomized interfaces (29) and the triplet is also strongly over-represented in trans-

membrane sequences (28). Thus, the frequent occurrence of Ser and Thr residues in transmembrane helices could be in part linked to their ability to engage in  $C\alpha-H\cdots O$  hydrogen bonds.

**Concluding Remarks.** By using known structures of helical membrane proteins, we have found common features in a number of helix-helix interfaces: (i) networks of apparent  $C\alpha-H\cdots O$  bonds; (ii) abundant interfacial Gly, Ser, and Thr residues; and (iii) short interhelical axial distances. In particular, Gly residues permit backbone-to-backbone  $C\alpha-H\cdots O=C$  formation by allowing short interhelical axial distances; the two  $\alpha$ -hydrogen atoms of Gly also increase the opportunities for  $C\alpha-H\cdots O$  formation. Ser and Thr side-chain hydroxyl groups allow  $C\alpha-H\cdots O$  interaction at longer interhelical distances. In addition,  $C\alpha-H\cdots O$  contacts appear more frequently in parallel right-handed helix-helix interactions. Finally, right-handed parallel GpA-like motifs, in which GxxxG pairs promote extended  $C\alpha-H\cdots O$  network formation, are a recurrent theme.

$C\alpha$  and carbonyl O atoms are typically shielded by side chains in an  $\alpha$ -helix, and so tend to be inaccessible. Derewenda *et al.* (14) defined  $C-H\cdots O$  contacts between helices as “esoteric” in soluble proteins and, in fact, they did not report a single instance of helical  $C\alpha-H\cdots O$  contacts with  $d_H < 2.7 \text{ \AA}$  in 13 high-resolution structures (1–2  $\text{\AA}$ ). Conversely, we have found 51 such contacts in 11 membrane proteins. Such a discrepancy is not due simply to the lower resolution of our database; instead, it depends on the frequency of Gly residues, which are rare in helical segments of soluble proteins but prevalent in transmembrane helices. It seems highly probable that  $C\alpha-H\cdots O$  hydrogen bonds exist in helix-helix interfaces having numerous Gly residues, very short interaxial distances, and backbone contacts.

Although  $C\alpha-H\cdots O$  hydrogen bonding could be merely incidental to the optimal packing achieved with close-range interhelical contacts, the key is then the relative energy contribution of the  $C\alpha-H\cdots O$  bond. If the estimate of 2.5–3.0 kcal/mol of interaction energy for the  $C\alpha-H\cdots O$  hydrogen bond is even approximately correct, several coordinated interactions would approximate or even exceed the energy of the  $N-H\cdots O$  hydrogen bond, which is known to drive nonspecific transmembrane helix association (10–13). However, packing would still be central in helix-helix interactions promoted by coordinated  $C\alpha-H\cdots O$  hydrogen bonds because main-chain atoms come in contact only between helices that fit together well. Moreover, because  $C\alpha-H\cdots O$  hydrogen bonds are weaker than main-chain  $N-H\cdots O=C$ , helix distortion would represent a substantial cost for their optimization. Thus, even if the energy contribution of backbone-to-backbone  $C\alpha-H\cdots O=C$  interactions turned out to be overwhelming with respect to the dispersion component of packing, their formation would be much more dependent on the specifics of local geometry than that of  $O-H\cdots O$  or  $N-H\cdots O$  bonds between ends of flexible side chains, thus reducing the dangers of nonspecific association. The  $C\alpha-H\cdots O$  could then be a more controllable and cooperative alternative to  $O-H\cdots O$  or  $N-H\cdots O$  bonds for exploiting the strength and directionality of hydrogen bonds in the hydrophobic environment and achieving, simultaneously, stability and specificity in transmembrane helix-helix interactions.

We thank J. Cabral, M. Cocco, R. Curran, J. Dawson, K. Ho, A. Lee, H. Li, M. Lemmon, N. Luscombe, K. MacKenzie, J.-L. Popot, J. Qian, B. Russ, and F. Zhou for helpful discussions and critical reading of the manuscript. This work was supported by grants from the National Institutes of Health and the National Science Foundation.

- Hendsch, Z. S. & Tidor, B. (1994) *Protein Sci.* **3**, 211–226.
- Hendsch, Z. S. & Tidor, B. (1999) *Protein Sci.* **8**, 1381–1392.
- Sheinerman, F. B., Norel, R. & Honig, B. (2000) *Curr. Opin. Struct. Biol.* **10**, 153–159.
- Popot, J. L. & Engelman, D. M. (1990) *Biochemistry* **29**, 4031–4037.
- Deisenhofer, J., Epp, O., Sinning, I. & Michel, H. (1995) *J. Mol. Biol.* **246**, 429–457.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., *et al.* (2000) *Science* **289**, 739–745.
- Popot, J. L. & Engelman, D. M. (2000) *Annu. Rev. Biochem.* **69**, 881–922.
- Ubarretxena-Belandia, I. & Engelman, D. M. (2001) *Curr. Opin. Struct. Biol.* **11**, 370–376.
- Snijder, H. J., Ubarretxena-Belandia, I., Blaauw, M., Kalk, K. H., Verheij, H. M., Egmond, M. R., Dekker, N. & Dijkstra, B. W. (1999) *Nature (London)* **401**, 717–721.
- Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T. & Engelman, D. M. (2000) *Nat. Struct. Biol.* **7**, 154–160.
- Zhou, F. X., Merianos, H. J., Brunger, A. T. & Engelman, D. M. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 2250–2255. (First Published February 13, 2001; 10.1073/pnas.041593698)
- Choma, C., Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2000) *Nat. Struct. Biol.* **7**, 161–166.
- Gratkowski, H., Lear, J. D. & DeGrado, W. F. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 880–885.
- Derewenda, Z. S., Lee, L. & Derewenda, U. (1995) *J. Mol. Biol.* **252**, 248–262.
- Bella, J. & Berman, H. M. (1996) *J. Mol. Biol.* **264**, 734–742.
- Fabiola, G. F., Krishnaswamy, S., Nagarajan, V. & Pattabhi, V. (1997) *Acta Crystallogr. D* **53**, 316–320.
- Desiraju, G. R. & Steiner, T. (1999) *The Weak Hydrogen Bond in Structural Chemistry and Biology* (Oxford Univ. Press, Oxford).
- Vargas, R., Garza, J., Dixon, D. A. & Hay, B. P. (2000) *J. Am. Chem. Soc.* **122**, 4750–4755.
- Scheiner, S., Kar, T. & Gu, Y. (2001) *J. Biol. Chem.* **276**, 9832–9837.
- Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (1999) *J. Mol. Biol.* **285**, 1735–1747.
- Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
- Sugeta, H. & Miyazawa, T. (1967) *Biopolymers* **5**, 673–679.
- Bansal, M., Kumar, S. & Velavan, R. (2000) *J. Biomol. Struct. Dyn.* **17**, 811–819.
- Bowie, J. U. (1997) *J. Mol. Biol.* **272**, 780–789.
- Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000) *Science* **290**, 481–486.
- Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000) *Nature (London)* **405**, 647–655.
- MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997) *Science* **276**, 131–133.
- Senes, A., Gerstein, M. & Engelman, D. M. (2000) *J. Mol. Biol.* **296**, 921–936.
- Russ, W. P. & Engelman, D. M. (2000) *J. Mol. Biol.* **296**, 911–919.
- Brosig, B. & Langosch, D. (1998) *Protein Sci.* **7**, 1052–1056.
- Font, M., Feliubadalo, L., Estivill, X., Nunes, V. V., Golomb, E., Kreiss, Y., Pras, E., Bisceglia, L., d’Adamo, A. P., Zelante, L., *et al.* (2001) *Hum. Mol. Genet.* **10**, 305–316.
- Lemmon, M. A., Treutlein, H. R., Adams, P. D., Brunger, A. T. & Engelman, D. M. (1994) *Nat. Struct. Biol.* **1**, 157–163.
- Op De Beeck, A., Montserret, R., Duvet, S., Cocquerel, L., Cacan, R., Barberot, B., Le Maire, M., Penin, F. & Dubuisson, J. (2000) *J. Biol. Chem.* **275**, 31428–31437.
- McClain, M. S., Cao, P. & Cover, T. L. (2001) *Infect. Immun.* **69**, 1181–1184.
- Wang, C. & Deber, C. M. (2000) *J. Biol. Chem.* **275**, 16155–16159.
- Javadpour, M. M., Eilers, M., Groesbeck, M. & Smith, S. O. (1999) *Biophys. J.* **77**, 1609–1618.
- Gray, T. M. & Matthews, B. W. (1984) *J. Mol. Biol.* **175**, 75–81.
- Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. & Pardo, L. (2000) *Biophys. J.* **79**, 2754–2760.