Empirical Lipid Propensities of Amino Acid Residues in Multispan Alpha Helical Membrane Proteins

Larisa Adamian,[1] Vikas Nanda,[2] William F. DeGrado,[2*] and Jie Liang[1*]

[1]*Department of Bioengineering, University of Illinois at Chicago, Illinois*
[2]*Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania*

ABSTRACT Characterizing the interactions between amino acid residues and lipid molecules is important for understanding the assembly of transmembrane helices and for studying membrane protein folding. In this study we develop TMLIP (<u>T</u>rans-<u>M</u>embrane helix-<u>LIP</u>id), an empirically derived propensity of individual residue types to face lipid membrane based on statistical analysis of high-resolution structures of membrane proteins. Lipid accessibilities of amino acid residues within the transmembrane (TM) region of 29 structures of helical membrane proteins are studied with a spherical probe of radius of 1.9 Å. Our results show that there are characteristic preferences for residues to face the headgroup region and the hydrocarbon core region of lipid membrane. Amino acid residues Lys, Arg, Trp, Phe, and Leu are often found exposed at the headgroup regions of the membrane, where they have high propensity to face phospholipid headgroups and glycerol backbones. In the hydrocarbon core region, the strongest preference for interacting with lipids is observed for Ile, Leu, Phe and Val. Small and polar amino acid residues are usually buried inside helical bundles and are strongly lipophobic. There is a strong correlation between various hydrophobicity scales and the propensity of a given residue to face the lipids in the hydrocarbon region of the bilayer. Our data suggest a possibly significant contribution of the lipophobic effect to the folding of membrane proteins. This study shows that membrane proteins have exceedingly apolar exteriors rather than highly polar interiors. Prediction of lipid-facing surfaces of boundary helices using TMLIP1 results in a 54% accuracy, which is significantly better than random (25% accuracy). We also compare performance of TMLIP with another lipid propensity scale, kPROT, and with several hydrophobicity scales using hydrophobic moment analysis. Proteins 2005;59:496–509.
© 2005 Wiley-Liss, Inc.

INTRODUCTION

The folding of helical membrane proteins involves the burial of residues from transmembrane (TM) helices in a lipid bilayer and the assembly of TM helices.[1–3] The contribution from side chains interacting with their environment reflects the energetic cost or gain due to the exposure of the residue to the lipid bilayer, or to the burial of the residue within the protein core. The contribution from interhelical interactions reflects the energetic cost or gain of various types of two-body and many-body interactions between transmembrane helices. The entropic effects include, among other terms, the restriction of the conformations of connected backbones and side chains. Quantitative estimation of these contributions is essential for model studies of membrane protein folding.

Here we estimate the free-energy cost or gain associated with the burial or exposure of different amino acid residue types to the lipid bilayer environment. This is an important endeavor, both for understanding the features stabilizing membrane proteins as well as the prediction of membrane protein structures. For example, an energetic scale for lipid exposure could be useful in differentiating properly folded from mis-folded structures generated by either *ab initio* or threading approaches to membrane protein structure prediction. Different lipid-contacting regions of a TM helix face either the highly hydrophobic hydrocarbon core or the more polar headgroup region of the lipid bilayer. Thus, different types of amino acid residues are likely to have different propensities for exposure to lipids at distinct regions of the helix–lipid interfaces. Indeed, Spencer and Rees[4] showed that helical TM proteins exhibit a central 20 Å-wide region with greater than 90% of its surface area contributed by carbon atoms, and very few formally charged atoms. On either side of this region the polarity of the protein increases in an approximately linear manner, reaching the distribution observed in water-soluble proteins after another 10 Å has been traversed.

Lipid propensity scales can be used for prediction of angular orientation, i.e., the helix rotation that decides buried versus exposed faces of TM helices. The transfer free energy of amino acid residues from solution to different regions of the lipid bilayer has been the focus of several experimental studies[5–7] that resulted in the development of the White-Wimley (WW) hydrophobicity scale, which is widely used for the prediction of TM helices in integral membrane proteins.[8] There are several other well-known hydrophobicity scales that have been used for the prediction of TM helices in membrane proteins: Kyte-Doolittle (KD),[9] Eisenberg-Weiss,[10] Goldman-Engelman-Steitz,[11] von Heijne,[12] Rost et al.[13,14] Using hydrophobicity moment calculations,[10,15] Rees and Eisenberg[16] showed that the average hydrophobicity of the TM surface of membrane proteins is higher than the average hydrophobicity of the interior of soluble proteins. However, the difference in the hydrophobicity of buried and exposed residues is smaller in membrane proteins than in water-soluble proteins, and therefore this approach has limited success in predicting the angular orientation of TM helices.[17] To improve the sensitivity of prediction of helical orientation, empirical lipid propensity scales were derived from analysis of membrane protein sequences[18,19] or structures.[20] Samatey et al.[18] used the periodic distribution of residues in the sequences of putative TM α-helices to extract a scale that describes the propensity of different amino acid residue types to lie on the buried or exposed faces of a TM helix. This scale is limited to the central part of TM helices that faces the hydrocarbon core of the phospholipid bilayer. Another empirical lipid propensity scale has been developed by Pilpel et al. from database analysis of membrane protein sequences that takes into account the different physico–chemical properties of the phospholipid bilayer and at the same time controls for the expected occurrence of residues in a null model.[19] Calculations using this scale showed promising results in predicting lipid-facing surfaces of structures of membrane proteins.

We develop in this study a lipid propensity scale of amino acid residues TMLIP (TM helix-LIPid) that measures their tendencies to partition into the core of the protein versus being exposed to lipids based on statistical analysis of a database of structures of multispan helical membrane proteins. We calculate lipid propensities separately for amino acid residue types in the headgroup region (TMLIP-H) and in the hydrocarbon core region (TMLIP-C). These parameters help to answer important questions about membrane proteins, for example, how does the burial of residues in a lipid bilayer differ from their burial in the interior of a protein? How different are the lipid propensities for the same residues located in different regions of the bilayer? Our study also helps to resolve the controversy concerning whether membrane proteins can be regarded as "inside-out" soluble proteins.[16,17,21] In addition, we use TMLIP lipid propensity scales to identify lipid-facing surfaces of multispan membrane proteins, given that a helix is known to be at the protein–lipid interface. This paper is organized as follows: we first describe the dataset and the computational meth-
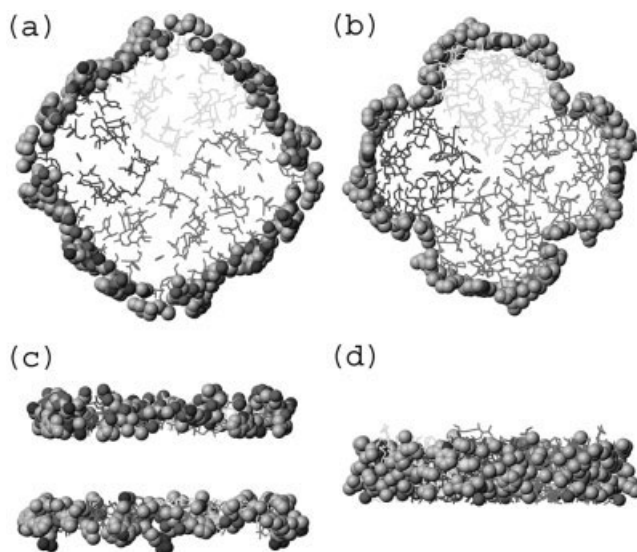

Fig. 1. Atoms on the surface of aquaporin tetramer (1J4N) that are accessible to 1.9 Å probe (shown in spacefill representation). **a**: Top view of headgroup region. **b**: Top view of hydrocarbon core region. **c**: Side view of headgroup region. **d**: Side view of hydrocarbon region.

ods. We then discuss the results and compare TMLIP with several hydrophobicity scales, followed by a discussion on predicting lipid facing surfaces of transmembrane helices.

METHODS

Lipid Probe Size

Protein–lipid interactions have been studied extensively by the application of the technique of electron spin resonance (ESR) spectroscopy using spin-labeled phospholipids.[22] ESR spectra showed the presence of a subpopulation of immobilized spin labels that are not observed in protein-free membrane. The interaction energy between the bound phospholipids and membrane proteins has a broad range. Some phospholipids transiently interact with a membrane protein, while others bind tightly to the grooves on its surface. To effectively sample the residues that interact with phospholipids, we probe the surface of TM helices with a sphere of radius 1.9 Å, which is the rounded up value of the effective van der Waals radius of a $-CH_2-$ group.[23] The advantage of using this sphere size is that it is small enough to probe the grooved surfaces of membrane proteins, but large enough to have a decreased access to more occluded residues of the protein in comparison with a traditional 1.4 Å probe. Figure 1 shows in spacefilling a diagram of the atoms on the TM surface of aquaporin tetramer (1J4N) that are accessible to a 1.9 Å probe. Panels (a) and (b) show top views of the headgroup and hydrocarbon core regions, respectively, while panels (c) and (d) show side views of the same regions.

Transmembrane Helices

The statistical analysis of phospholipid-facing residues is based on a set of 29 alpha-helical TM proteins (Table I). Protein structures in their native quaternary state were used when available (e.g., trimer for bacteriorhodopsin

TABLE I. Dataset of Structures of Membrane Proteins Used in this Study

PDB	Protein	Origin	Å	Oligomer
1. 1C3W	Bacteriorhodopsin	*H. salinarum*	1.6	Trimer
2. 1E12	Halorhodopsin	*H. salinarum*	2.4	Trimer
3. 1EHK	Ba3 cytochrome c oxidase	*T. termophilus*	2.4	Monomer
4. 1EUL	Calcium ATPase	*O. cuniculus*	2.6	Monomer
5. 1FX8	Glycerol facilitator	*E. coli*	2.2	Tetramer
6. 1H2S	Sensory rhodopsin II	*N. pharaonis*	1.9	Dimer
7. 1IWG	Multidrug efflux transporter	*E. coli*	3.5	Trimer
8. 1J4N	Aqp1 water channel	*B. Taurus*	2.2	Tetramer
9. 1K4C	Potassium channel Kcsa	*S. lividans*	2.0	Tetramer
10. 1KB9	Cytochrome bcl complex	*S. cerevisiae*	2.3	Dimer
11. 1KF6	Quinol-fumarate reductase	*E. coli*	2.7	Monomer
12. 1KPL	Clc chloride channel	*S. typhimurium*	3.0	Dimer
13. 1KQF	Formate dehydrogenase N	*E. coli*	1.6	Trimer
14. 1L7V	Vitamin B12 transporter	*E. coli*	3.2	Dimer
15. 1L9H	Rhodopsin	*B. taurus*	2.6	Monomer
16. 1M3X	Photosynthetic reaction center	*R. sphaeroides*	2.6	Monomer
17. 1M56	Cytochrome c oxidase	*R. sphaeroides*	2.3	Monomer
18. 1MSL	Mechanosensitive channel	*M. tuberculosis*	3.5	Pentamer
19. 1NEK	Succinate dehydrogenase	*E. coli*	2.6	Trimer
20. 1OCR	Cytochrome c oxidase	*B. Taurus*	2.3	Dimer
21. 1OKC	ADP/ATP carrier	*B. Taurus*	2.2	Dimer
22. 1PP9	Cytochrome bcl complex	*B. Taurus*	2.1	Dimer
23. 1PV6	Lactose permease	*E. coli*	3.6	Monomer
24. 1PW4	Glycerol-3-phosphate transporter	*E. coli*	3.3	Monomer
25. 1Q16	Nitrate reductase A	*E. coli*	1.9	Dimer
26. 1QLA	Fumarate reductase	*W. succinogenes*	2.2	Dimer
27. 1RC2	Aquaporin Z	*E. coli*	2.5	Tetramer
28. 1RH5	Protein conducting channel	*M. jannaschii*	3.2	Monomer
29. 1UM3	Cytochrome b_6f complex	*M. laminosus*	3.0	Monomer

(1C3W), and tetramer for aquaporin (1J4N)). In this approach, helix–helix interfaces of the oligomers were treated as lipid-inaccessible surfaces, although these helices would all face lipid if the structures existed as monomers. Transmembrane helices were determined visually. Heme and other covalently bound cofactors are kept in protein structures (e.g., photosynthetic reaction center (1M3X), cytochrome bc1 complex (1PP9, 1KB9) and cytochrome c oxidase (1OCR)), because they shield TM helices from interactions with phospholipids.

The phospholipid bilayer has two chemically distinct regions: the hydrocarbon core region and headgroup region. The combined thickness of both headgroup regions is approximately equal to the thickness of the hydrocarbon core.[3] Therefore, each TM helix is divided into four quarters: the two outer quarters are considered as the headgroup region, and the two inner quarters as the hydrocarbon core region. We estimate lipid propensities of amino acid residues in each region separately. For calculation of residue-based lipid propensities, we classify residues located at the borders of core and headgroup regions by the location of their C_α atoms. This approach does not account for the "snorkeling" effects of side chains of Lys and Arg. The polar side chains of a snorkeling Lys or Arg may be located in the headgroup region. However, if its C_α atom is located in the hydrocarbon core region, the residue is classified as a core residue. The same approach is applied to other residues with large side chains (Trp and Tyr). The

advantage is that structure-derived propensity scales can be readily applied directly to the amino acid sequence of the TM helix, as the specific region a residue belongs to will only depend on its position in the TM helix.

Calculation of Probe-Accessible Amino Acid Residues

We use the VOLBL (www.alphashapes.org/alpha/read-mebuvo.html) method to compute probe-accessible residues. VOLBL uses precomputed Delaunay triangulation and alpha shape to measure metric properties of protein structures. The Delaunay triangulation of a membrane protein is computed using the DELCX program,[24,25] and the alpha shape is computed using the MKALF program.[24,26] All programs can be downloaded from the website of the National Center for Supercomputing and its Applications (http://www.ncsa.uiuc.edu). The van der Waals radii of protein atoms are taken from Tsai et al.[23] Exposed residues have > 0.0 Å^2 solvent accessibility.

Lipid Propensity

The propensity P_i of an individual residue type i to interact with phospholipids is defined as the ratio of the probability $\pi_{i,s}$ of being lipid accessible to the probability π_i of being buried:

$$P_i = \pi_{i,s} / \pi_i, \quad (1)$$

where

$$\pi_{i,s} = n_{i,s} \,/\, n_s, \tag{2}$$

$$\pi_i = n_i \,/\, n. \tag{3}$$

For a specific region (hydrocarbon core or headgroup region), $n_{i,s}$ is the number of probe-accessible (surface) residues of type i. We calculated two scales using different references states: for TMLIP1, n_s is the *total* number of probe accessible residues, n_i is the *total* number of residues of type i in the region, and n is a total number of residues in the region, for TMLIP2, n_i is the number of *buried* residues of type i, and n is a total number of *buried* residues in the region. We list the logarithmic value $\ln P_i$. Residues with $\ln P_i > 0$ have a tendency to face lipids, and residues with $\ln P_i < 0$ tend to face away from lipid. We follow our earlier study[27] and use 1,000 resamplings of bootstrap data to calculate the 95% confidence intervals for the estimated propensities.

Our approach is different from that of Beuming and Weinstein,[20] where propensity is defined as average fractions of exposed surface area of a residue type after normalization by a constant. Our approach employs an explicit null model, i.e., a reference state: namely, the random probability of finding a residue type in a specific region.

Calculation of Helical Lipophilicity Moments

To evaluate objectively the effectiveness of estimated lipid propensities, we assess the accuracy of prediction of lipid-facing surfaces for helices known to be at the protein–lipid boundary by calculating helical lipophilicity moments. With the availability of high-resolution coordinates for membrane proteins, it is no longer necessary to make approximations of fixed periodicity using Fourier transform as was done in previous studies.[28,29] Instead, i and j components of each helix moment are calculated using x and y coordinates of a helix aligned along the z-axis. The moment is calculated using the following expression:

$$\mathbf{M} = \bar{\imath} \cdot \sum_{n=1}^{N} U_n \cdot x_n + \bar{\jmath} \cdot \sum_{n=1}^{N} U_n \cdot y_n, \tag{4}$$

where U_n is the property (accessible area or propensity) for the nth residue in the helix, x_n and y_n are x and y coordinates for the C_α atom of the nth residue. Lipophilicity moments are calculated by applying headgroup TMLIP and kPROT propensities for the first and the last quarter of TM residues, and hydrocarbon core TMLIP and kPROT propensities to the middle two quarters of the helix. When calculating moments using other hydrophobicity scales, only the core regions of the TM helices are used, as hydrophobicity scales are not applicable for residues in the headgroup region. We find that the calculated lipophilicity moment depends on the definition of the exact boundary between interface and hydrocarbon regions of the TM helix. Solvent accessibility of every amino acid residue type X is calculated as a fraction relative to a helical reference state defined as an idealized α-helix with 3.6 residues per turn and the sequences $(Gly)_4$-X-$(Gly)_4$. Solvent accessibility moment is calculated similarly by Equation (4) with U_n being the probe accessible surface area.

RESULTS AND DISCUSSION
TMLIP1 Propensities for Amino Acid Residue Types to Interact with Phospholipids

We first calculate a lipid propensity scale using all residues in the respective region of a structure as a reference state. This lipid propensity scale is called TMLIP1. In order to make a comparison to other hydrophobicity scales, we follow the definition of Chothia et al.[30] and use buried residues as a reference state to calculate a second lipid propensity scale, which we call TMLIP2. TMLIP1 scale is used for prediction of lipid-exposed surfaces of TM helices, while TMLIP2 scale is used for comparison of various transfer free energies. The estimated residue lipid propensities ($\ln P_i$) of TMLIP1 for each type i amino acid residue in each region are shown in Table II, along with the 95% confidence intervals, and the total number of accessible and buried residues of each type observed in the full data set. To account for the symmetry formed by the repeating monomeric subunits in the oligomeric structures, we divide the total number of lipid-exposed or buried residues in the complex by the number of monomers in the oligomer.

Headgroup region

Polar and ionizable amino acid residues such as Lys (estimated residue log lipid propensity 0.24) and Arg (0.14) have a tendency to be exposed to phospholipids. They are likely to participate in direct or water-mediated polar–polar interactions with phospholipid headgroups or the glycerol backbone. Aromatic residues Trp (0.25) and Phe (0.13) also have a strong propensity to face phospholipids in the headgroup region. Aromatic residues, especially Trp, are thought to act as anchors for a membrane protein.[31] The π electron structure and the electronic quadrupole moment associated with Trp favors location in the headgroup region.[32] The corresponding value for Tyr (0.06) has a confidence level $(-0.03, 0.10)$ that spans both the favorable region ($\ln P > 0$) and unfavorable region ($\ln P < 0$).

Among the large aliphatic residues such as Ile (0.06), Leu (0.09), Met (-0.03) and Val (-0.03), only Leu has a significant tendency to face lipids in the headgroup region, while other residues have no preference to be buried inside of the TM bundle or to be exposed to phospholipid molecules. Small residues and Thr show a strong preference for being buried within the TM helical bundle, regardless of the region of the membrane: Ala (TMLIP1-H: -0.12, TMLIP1-C: -0.06), Gly $(-0.34, -0.48)$, Ser $(-0.22, -0.29)$, Thr $(-0.15, -0.16)$. This is consistent with the observation that small residues such as Gly and Ala have strong propensities for interhelical interactions.[27,33,34] His is the only strongly polar residue that has a stronger tendency to be buried in the protein interior within the headgroup region. The remaining polar residues (Asn, Asp, Gln, and Glu) fail to show a statistically significant bias to be either buried or exposed in this region of the membrane.

TABLE II. TMLIP1 and TMLIP2 Lipid Propensities and Transfer Energies for Headgroup and Hydrocarbon Core Regions of TM Helices[†]

A. Headgroup region

	Accessible $N_{h,a}$	Unacc. $N_{h,b}$	TMLIP1-H propensity		TMLIP2-H propensity		Transfer energy	
			$\ln P_{i,h}$	Bootstrap interval	$\ln P_{i,h}$	Bootstrap interval	ΔG, kcal/mol	Bootstrap interval
ALA	159	139	−0.12	−0.19 . . . −0.04	−0.34	−0.58 . . . 0.12	0.20	0.35 . . . 0.07
ARG	93	52	0.14	−0.02 . . . 0.20	0.39	−0.07 . . . 0.82	−0.23	0.04 . . . −0.49
ASN	42	36	0.07	−0.14 . . . 0.17	−0.05	−0.56 . . . 0.38	0.03	0.34 . . . −0.23
ASP	39	26	0.04	−0.20 . . . 0.15	0.18	−0.31 . . . 0.67	−0.11	0.19 . . . −0.40
CYS	15	13	−0.12	−0.80 . . . 0.07	0.14	−0.97 . . . 0.95	−0.09	0.58 . . . −0.57
GLN	41	45	−0.08	−0.22 . . . 0.10	−0.50	−1.11 . . . 0.06	0.30	0.66 . . . 0.04
GLU	46	28	−0.03	−0.26 . . . 0.08	−0.11	−0.73 . . . 0.39	0.06	0.44 . . . −0.23
GLY	103	140	−0.34	−0.46 . . . −0.19	−0.74	−1.05 . . . −0.43	0.44	0.63 . . . 0.26
HIS	49	41	−0.13	−0.65 . . . 0.02	−0.19	−0.94 . . . 0.34	0.12	0.56 . . . −0.20
ILE	137	92	0.06	−0.04 . . . 0.10	0.11	−0.17 . . . 0.39	−0.06	0.10 . . . −0.23
LEU	281	159	0.09	0.05 . . . 0.17	0.17	−0.11 . . . 0.43	−0.10	0.06 . . . −0.26
LYS	82	29	0.24	0.12 . . . 0.31	0.87	0.41 . . . 1.30	−0.52	−0.25 . . . −0.77
MET	85	52	−0.03	−0.12 . . . 0.08	0.32	−0.16 . . . 0.79	−0.19	0.10 . . . −0.47
PHE	175	85	0.13	0.10 . . . 0.23	0.45	0.14 . . . 0.77	−0.27	−0.08 . . . −0.46
PRO	60	45	−0.04	−0.14 . . . 0.09	−0.01	−0.46 . . . 0.41	0.01	0.28 . . . −0.25
SER	73	79	−0.22	−0.30 . . . −0.12	−0.37	−0.67 . . . −0.05	0.22	0.40 . . . 0.03
THR	85	85	−0.15	−0.30 . . . −0.04	−0.33	−0.69 . . . 0.02	0.20	0.41 . . . −0.01
TRP	113	39	0.25	0.14 . . . 0.31	0.90	0.40 . . . 1.40	−0.53	−0.24 . . . −0.83
TYR	56	55	0.06	−0.03 . . . 0.10	0.21	−0.07 . . . 0.49	−0.12	0.04 . . . −0.29
VAL	143	105	−0.03	−0.11 . . . 0.07	−0.02	−0.36 . . . 0.29	0.01	0.21 . . . −0.17

B. Hydrocarbon region

	Accessible $N_{h,a}$	Unacc. $N_{h,b}$	TMLIP1-C propensity		TMLIP2-C propensity		Transfer energy	
			$\ln P_{i,h}$	Bootstrap interval	$\ln P_{i,h}$	Bootstrap interval	ΔG, kcal/mol	Bootstrap interval
ALA	255	247	−0.06	−0.14 . . . 0.04	−0.20	−0.42 . . . 0.02	0.12	0.25 . . . −0.01
ARG	6	18	−1.47	−4.61 . . . −0.40	−1.73	−3.91 . . . −0.78	1.03	2.33 . . . 0.46
ASN	12	39	−1.20	−1.83 . . . −0.80	−1.68	−2.53 . . . −1.14	1.00	1.51 . . . 0.68
ASP	8	16	−0.87	−1.83 . . . −0.17	−1.22	−2.81 . . . −0.19	0.73	1.68 . . . 0.11
CYS	28	25	−0.15	−0.31 . . . 0.06	0.01	−0.63 . . . 0.52	−0.01	0.38 . . . −0.31
GLN	8	24	−0.94	−2.30 . . . −0.65	−1.33	−2.41 . . . −0.63	0.79	1.44 . . . 0.38
GLU	7	23	−0.69	−1.11 . . . −0.16	−1.65	−3.91 . . . −0.76	0.98	2.33 . . . 0.45
GLY	127	264	−0.48	−0.62 . . . −0.37	−1.05	−1.27 . . . −0.82	0.62	0.76 . . . 0.49
HIS	8	45	−1.02	−1.77 . . . −0.71	−2.12	−3.91 . . . −1.39	1.26	2.33 . . . 0.83
ILE	316	132	0.19	0.13 . . . 0.25	0.63	0.41 . . . 0.83	−0.38	0.25 . . . −0.50
LEU	484	205	0.17	0.15 . . . 0.24	0.54	0.34 . . . 0.71	−0.32	0.20 . . . −0.42
LYS	9	16	−0.39	−0.78 . . . −0.02	−0.86	−2.12 . . . −0.12	0.51	1.26 . . . 0.07
MET	98	76	0.03	−0.13 . . . 0.10	0.02	−0.34 . . . 0.31	−0.01	0.20 . . . −0.19
PHE	281	120	0.24	0.13 . . . 0.29	0.68	0.36 . . . 1.03	−0.40	−0.21 . . . −0.61
PRO	48	52	−0.16	−0.27 . . . 0.02	−0.43	−0.82 . . . −0.13	0.26	0.49 . . . 0.08
SER	73	113	−0.29	−0.49 . . . −0.20	−0.59	−0.92 . . . −0.30	0.35	0.55 . . . 0.18
THR	114	115	−0.16	−0.24 . . . −0.07	−0.42	−0.73 . . . −0.20	0.25	0.44 . . . 0.12
TRP	69	27	0.06	−0.01 . . . 0.22	0.50	−0.17 . . . 1.13	−0.30	0.10 . . . −0.68
TYR	56	51	−0.01	−0.33 . . . 0.11	−0.06	−0.58 . . . 0.43	0.04	0.35 . . . −0.25
VAL	335	155	0.21	0.11 . . . 0.22	0.57	0.34 . . . 0.81	−0.34	−0.20 . . . −0.48

[†]$\ln P_{i,h}$ and $\ln P_{i,c}$ estimated log value of residue propensities for the head group region (h) and hydrocarbon core region (c), respectively; Transfer energy: $\Delta G = -RT\ln P_{i,h}$ and $\Delta G = -RT\ln P_{i,c}$ in kcal/mol at 27°C. Bootstrap intervals: the lower end and upper end of 95% confidence intervals estimated by 1,000 bootstraps; $N_{h,a}$, $N_{h,b}$ and $N_{c,a}$ $N_{c,b}$: Number of accessible and buried residues in headgroup and hydrocarbon core regions, respectively. Residues with negative values of $\ln P_i$ tend to face away from lipid, while those with positive value of $\ln P_i$ tend to face toward lipid.

Hydrocarbon core region

The strongest preference for interacting with lipids is observed for Ile (TMLIP1-C = 0.19), Leu (0.17), Phe (0.24), and Val (0.21). These four residue types have the highest abundance and make up about 60% of all lipid-facing residues in the hydrocarbon region. Trp has an estimated propensity value of 0.06, showing some preference to face lipids in the hydrocarbon region. Structural analysis of Trp residues that are assigned to the hydrocarbon core region showed that most of them are located near the

headgroup region with NE1 atoms often pointing towards the ends of the TM helices. Large fractions of the Trp side chains are likely to be positioned physically in the headgroup region. Met is marginally more abundant on the surface of the hydrocarbon core region than the headgroup region (98 residues vs. 85), but it exhibits essentially the same neutral preference to be either buried or exposed. The Met side chain is flexible and can sample many conformations, which makes it versatile for interhelical packing. Analysis of higher order interhelical interactions in membrane proteins[27] showed many types of high propensity interhelical triplets that contain methionine.

Polar amino acid side chains are almost nonexistent on the surface of the hydrocarbon core region. They have high propensities to be buried inside of the TM bundle, where they often fulfill functional roles and promote helix–helix interactions through interhelical hydrogen bonding.[35–38] Moreover, polar interactions of asparagines are position-dependent in a membrane environment, as has been experimentally demonstrated in synthetic membrane-solubilized GCN4 oligomers,[39] transmembrane leucine zippers,[40] and the transmembrane region of the erythropoietin receptor.[41] These studies show that Asn side chains provide a significantly larger driving force for helix association within the hydrocarbon core region of the TM helix rather than near the headgroup regions.

Effect of Data Set of Homologous Structures on Propensity Scale

Residues on the surfaces of soluble as well as membrane proteins are less conserved than the residues buried within the core of the protein.[16,28,42,43] We expanded our data set by including two homologous structures of cytochrome c oxidase from *Bos taurus* (1OCR) and *Rhodobacter sphaeroides* (1M56) and two homologous structures of cytochrome bc1 complexes from *Saccharomyces cerevisiae* (1KB9) and *Bos taurus* (1PP9). The sequence identity between subunits I, II, and III of cytochrome c oxidase is 50%, 35%, and 48%, respectively. However, the bovine protein is composed of a larger number of subunits, which results in substantially different lipid-facing surfaces. In cytochrome bc1 complexes, C and D chains have 51% and 58% sequence identity, respectively.

To assess the effect of including homologous structures on the derived lipid propensity values, we run control calculations on a reduced data set consisting of 27 proteins with no more than 30% sequence identity between any pair of structures. Overall, the propensity values experience only a very small change. Small increases or decreases of the observed propensity values were observed for several amino acid types, the majority of which are not sufficiently sampled. For example, Asn (TMLIP1-H = 0.07 calculated from the set of 29 proteins versus −0.02 calculated from the set of 27 proteins), Glu (−0.03 vs. −0.14) and His (−0.13 vs. −0.39) in the interface and Asp (TMLIP1-C = −0.87 vs. −1.66), Gln (−0.94 vs. −1.14), Glu (−0.69 vs. −1.14) and Tyr (−0.01 vs. −0.11) in the hydrocarbon regions. Minimal changes observed for headgroup region propensity values for abundant amino acid residues such as Ala (−0.12 vs. −0.14), Gly (−0.34 vs. −0.34), Ile (0.06 vs. 0.05), Leu (0.09 vs. 0.10) and Phe (0.17 vs. 0.13). Abundant residues in the hydrocarbon group region behave similarly, e.g. TMLIP1-C for Ala (−0.06 vs. −0.11), Gly (−0.48 vs. −0.56), Ile (0.19 vs. 0.22), Leu (0.17 vs. 0.23), Phe (0.24 vs. 0.30) and Val (0.21 vs. 0.24). The confidence intervals obtained by bootstrap sampling are naturally tighter in the "full" data set composed of 29 proteins in comparison with the reduced data sets. For example, a confidence interval for Ala in the headgroup region is (−0.19 . . . −0.04) in the 29-protein data set, while it widens to (−0.20 . . . −0.02) in the 27-protein data set. Similarly, a confidence interval for Leu in the hydrocarbon core is smaller (0.15 . . . 0.24) in the 29-protein data set than in the 27-protein data set (0.17 . . . 0.29). Tighter bootstrap intervals for lipid propensity values are found for 16 amino acid types in the interface and for 13 amino acid types in the hydrocarbon regions. Comparison of estimated TMLIP propensity values and bootstrapped confidence intervals obtained with full and reduced data sets shows that the addition of two homologous structures increases the reliability of potentials by tightening bootstrap intervals without significant changes of observed lipid propensity values.

TMLIP2 and Hydrophobicity Scales

We can convert lipid propensities into free energies of transfer, which provide a measure of the favorability of transferring a given side chain from the lipid-inaccessible interior of a protein to the hydrocarbon region of the bilayer. If membrane proteins are more nonpolar on their exterior than their interior, one would expect that the transfer energy values will correlate with various amino acid hydrophobicity scales. Following Miller et al., we convert lipid propensity into a free energy of transfer through the expression: $\Delta G_t = -RT \ln P_i$, where RT = 0.596 kcal/mol at 27°C,[30] and $\ln P_i$ is the TMLIP2 propensity value for residue type i. We found that indeed there is a good correlation between TMLIP2 transfer energies in the hydrocarbon region and the scales of Chothia,[30] White and Wimley's free energy of transfer from water to octanol,[3] or that of Eisenberg and Weiss[44] [Fig. 2(a,c,e)].

Membrane proteins have exceedingly apolar exteriors

Chothia's scale was computed using water-soluble proteins and has a sign opposite to that of TMLIP2. The good correlation between the two scales in the hydrocarbon core [Fig. 2(a)] clearly shows that membrane proteins have more polar residues on their interiors than their exteriors, while the opposite is true for water-soluble proteins. This correlation also provides unambiguous support for the results of variability and hydrophobicity analysis of TM helices by Rees and Eisenberg,[16] who concluded that membrane and water-soluble proteins exhibit comparable interior characteristics and differ primarily in the chemical polarity of the surface residues, which are exceedingly apolar in membrane and polar in globular proteins. To further examine this issue, we
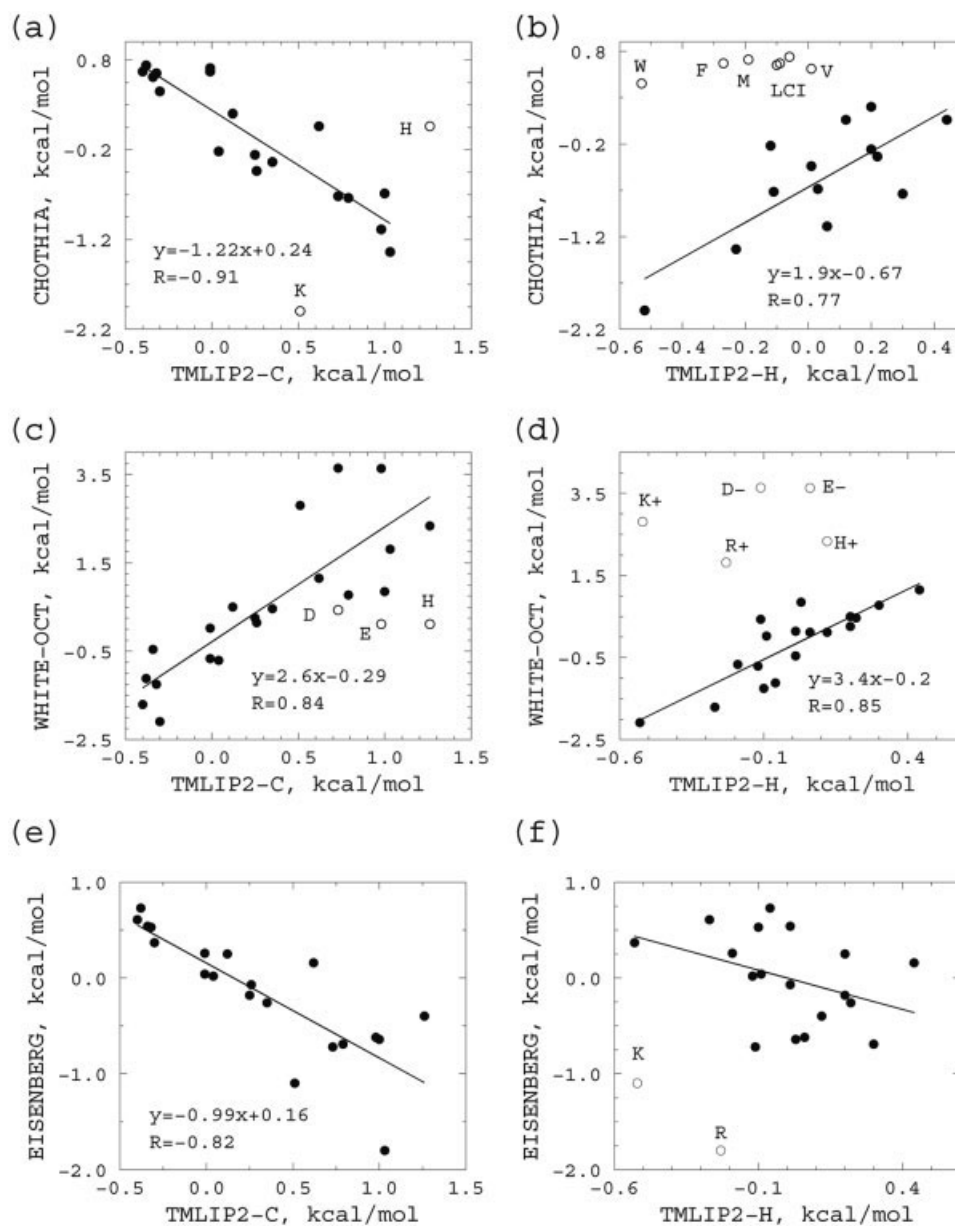
Fig. 2. Correlation between TMLIP2-C (panels a, c, e) and TMLIP2-H (panels b, d, f) free energies of transfer and hydrophobcity free energy scales. Filled symbols represent residues that were included in regression calculations. **a,b**: TMLIP2-C and TMLIP2-H versus the scale of Chothia.[30] A good correlation (R = 0.77) between TMLIP2-H and Chothia's scales for the subset of polar and charged residues is seen, where Phe, Met, Leu, Ile, Cys, and Val cluster together with Trp and have close values within the interval of 0.65–0.74 in Chothia's scale. **c,d**: TMLIP2-C and TMLIP2-H versus whole-residue hydrophobicity scale of transfer from water to octanol (White-Wimley[3]). Better correlation with TMLIP2-C was obtained when transfer energy values for His, Glu, and Asp were taken in their charged states and when Lys and Arg are excluded. Points corresponding to neutral states are shown with unfilled circles (panel c). In case of TMLIP2-H, a better correlation was obtained when transfer energy values for His, Glu and Asp were taken in their neutral states. Points corresponding to their charged states are shown as unfilled circles (panel d). **e,f**: TMLIP2-C and TMLIP2-H verus Eisenberg-Weiss[44] hydrophobicity scale.

used our data set of buried and exposed residues from membrane protein structures and calculated average hydrophobicities using the Eisenberg-Weiss scale as described by Rees and Eisenberg.[16] We found that the average hydrophobicity of the interior of 29 membrane proteins is 0.17 and the average hydrophobicity of the exterior is 0.38. The respective numbers obtained by Rees and Eisenberg[16] for 35 TM helices based on hydrophobicity moment calculations are 0.15 and 0.34. However, the same parameters for water-soluble oligomeric proteins obtained with the same method and application of the same hydrophobicity scale of Eisenberg-Weiss are 0.19 and −0.28.[16]
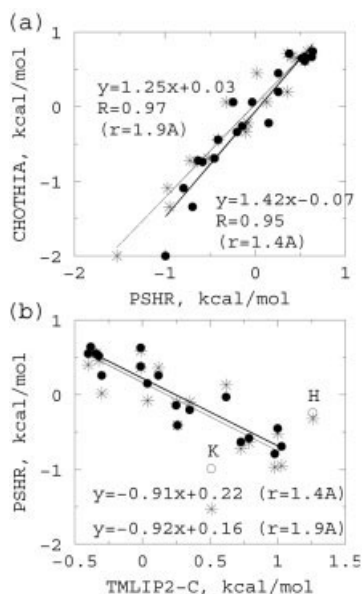
Fig. 3. **a**: Comparison of Chothia[30] and PSHR scales. PSHR scale is calculated from 622 high-resolution structures from PDB SELECT[45] database using α shape approach with probe radii 1.4 Å (filled circles and solid regression line, panels a, b) and 1.9 Å (stars and dashed regression line, panels a, b). **b**: Correlation between TMLIP2-C and PSHR scales. Correlation coefficients R for PSHR scale calculated with 1.4-Å and 1.9-Å probes are −0.92 and −0.87, respectively. Lys and His were excluded from the regression calculations.

TMLIP2 scale is determined in a manner similar to that of Chothia's scale,[30] but with a different method to calculate accessible surface area. To investigate the effect of different calculations of accessible surface area, we derived hydrophobic scales of soluble proteins using 622 X-ray structures from PDB SELECT data base,[45] where each structure has a resolution of 2.5 Å or better. We calculated accessible surface area using probe radius of 1.4 Å and 1.9 Å, respectively, and compared the results (called PSHR scale for "PDB SELECT with High Resolution") with Chothia's scale [Fig. 3(a)]. Both PSHR scales calculated using probe size of 1.4 Å and 1.9 Å showed strong correlations with Chothia's scale (correlations coefficients of 0.95 and 0.97, respectively). The values of transfer energies in Chothia's scale are larger than in both PSHR scales, reflected by slope values of 1.42 and 1.25 for 1.4 Å and 1.9 Å probes, respectively.

Lipophobic and hydrophobic forces are similar in magnitude

It is interesting to compare the relative strengths of the hydrophobic effect in water and the lipophobic effect in phospholipid bilayers, which are relevant to the folding of soluble and membrane proteins, respectively. The scales of Chothia, PSHR, and TMLIP2 are derived in a similar manner, which allows such a comparison. The regression slope on Figure 2(a) has a value of −1.2, the corresponding value for comparison of TMLIP2 scale with PSHR scales [Fig. 3(b)] is −0.9. These values are close to −1.0, indicating that the lipophobic and hydrophobic forces are similar

in magnitude, but opposite in sign. The lipophobic effect, as defined here, is a purely empirical property of the amino acid side chains. The lipophilicity of a given side chain would have contributions from its ability to preferentially interact in the protein interior versus the lipid-exposed surface. Polar residues are strongly lipophobic, presumably because they can form more stabilizing interactions in the buried core of a protein than when exposed to the fatty acyl chains of a membrane. Another contributor to the lipophobic effect arises from unfavorable loss of entropy of lipid molecules interacting with helices, instead of freely diffusing in the membrane bilayer. Experiments using electron spin resonance spectroscopy showed the presence of a subpopulation of immobilized lipids when proteins are present in membrane.[22] However, these phospholipid molecules are in fast exchange with the bulk phospholipids, suggesting such nonspecific membrane protein–lipid interactions are not favorable in the membrane. The entropic contribution to the lipophobic effect may provide a general force for the aggregation of TM helices, while the final structure largely depends on the more specific interhelical hydrogen bonding, van der Waals interactions, as well as constraints provided by the interhelical loops.

This result supports prediction made by Rees et al.[46] who proposed that the work associated with placing a protein in a solvent could be, to first order, independent of the solvent based on the similarities in surface areas between the photosynthetic reaction center and water-soluble proteins of similar size. The authors suggested that the surface energies of these proteins must be similar, despite of the differences in surface tensions between hydrocarbon liquids ($\sim$ 30 cal/Å^2) and water ($\sim$ 100 cal/Å^2), because the greater energy required to create a surface in water can be offset by the greater favorable interaction between the polar surface residues and water, while the interactions between nonpolar surface residues and the hydrocarbon chains in the bilayer are weaker. As a consequence of these compensating effects, the net result could be comparable surface energies for the interaction of the relevant solvents with either water-soluble or membrane proteins. Rees et al.[46] called this effect "solvophobic" and suggested that like hydrophobic effects, solvophobic effects will also tend to minimize the exposed surface area and create compactly folded structures.

Correlation Between Various Hydrophobicity Scales and TMLIP2

Table III provides the degree of correlation between various hydrophobicity scales and computed TMLIP2. The degree of correlation compares well with that between the individual hydrophobicity scales (0.64 for White-Wimley vs. Chothia; 0.94 for Chothia vs. Eisenberg, and 0.84 for White-Wimley vs. Eisenberg). For residues in the head-group region, the overall correlations are poor between TMLIP2 and other scales, but if various subset residues are examined, TMLIP2 shows good correlation with these hydrophobicity scales [Fig. 2(b,d,f); Table III].

TABLE III. Correlation of TMLIP2 and Various Hydrophobicity Scales[†]

	Chothia[30]	White-Wimley (Octanol-water)[3]	Weiss and Eisenberg[44]	White-Wimley (interface − water)[3]
TMLIP-C	−0.71	0.84	−0.82	0.71
TMLIP-C subset	−0.91 (K, H)	n/a D, E, H charged	n/a	0.74 (K, W)
TMLIP-H	0.07	0.20	0.06	0.37
TMLIP-H subset	0.77 (W, F, M, L, C, I, V)	0.85 (K, R) D, E, H neutral	−0.39 (K, R)	0.77 (K, R)

[†]Residues enclosed in parenthesis are excluded from the regression calculations.
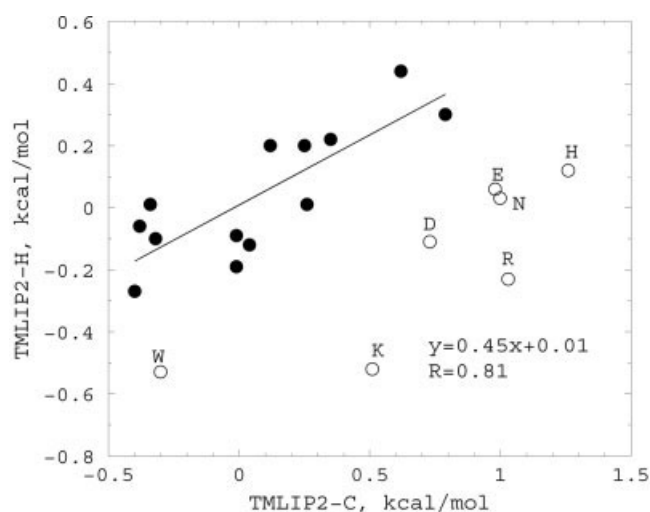

Fig. 4. Correlation of transmembrane lipid propensities for residues in the headgroup region (TMLIP2-H) and in the hydrocarbon core region (TMLIP2-C). Except Trp, all nonpolar and weakly polar amino acid residues have strongly correlated lipid propensities at these two regions. Correlation coefficient applies to the filled circles only.

Fig. 5. Correlation of transfer free energy of amino acid residues from core to headgroup region as reflected by TMLIP2 difference (ΔTMLIP2 = TMLIP2-C − TMLIP2-H) between these two regions and other hydrophobicity scales. Correlation of ΔTMLIP2 with: (**a**) White's[3] $\Delta\Delta G_t$; (**b**) Chothia's[30] scale, and (**c**). Eisenberg's[44] scale. All values are in kcal/mol.

Significant Correlations Between TMLIP2 for Nonpolar Residues in the Headgroup Region Versus the Hydrophobic Core Region

Figure 4 illustrates the correlation between the ΔG_t values of TMLIP2 for the core versus the headgroup region of the bilayer. An excellent correlation is observed for all the nonpolar (Ala, Val, Leu, Ile, Met, Phe) and weakly polar (Thr, Ser, Tyr, Gly, Cys, Pro) amino acids. The only exception is Trp, which is specifically stabilized in the headgroup region.[31,32,47] The slope of the line (excluding Trp, Fig. 3) is 0.45, which may indicate that the positional lipid propensities are attenuated in the headgroup region, presumably because this region has both hydrophobic as well as hydrophilic characters. The strong polar residues are also much more likely to be exposed in the headgroup region, presumably because they are better solvated in this region. Furthermore, they should be able to form specific interactions with polar functionality of the headgroups.

Transfer From the Hydrophobic Core Region to the Headgroup Region of the Bilayer

The difference between TMLIP2 transfer energy values for the interfacial headgroup region versus the hydrophobic core region of the bilayer should reflect ease of transfer of lipid-exposed residue from the bilayer center t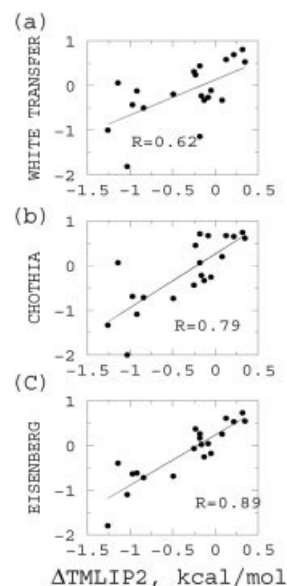o the headgroup region. We therefore computed ΔTMLIP2, which is the difference between the values for ΔG_{th} of the headgroup versus the ΔG_{tc} of the core regions, and compared it to various measures of hydrophobicity as well as the propensity of residues for the bilayer headgroup region. White and Wimley[3] have computed the difference between the free energy of transfer of amino acid side chains (designated here as $\Delta\Delta G_{tr}$) from octanol to water versus transfer from water to the interfacial region of the bilayer. The value of ΔTMLIP2 correlates with $\Delta\Delta G_{tr}$ similarly [correlation coefficient R = 0.62; Fig. 5(a)] with free energy of transfer from water to octanol (R = −0.63, data not shown), and from water to the interface region (R = −0.66, data not shown). However, ΔTMLIP2 shows an even better correlation with either the scale of Chothia [R = 0.78; Fig. 5(b)] or Eisenberg [R = 0.89; Fig. 5(c)], indicating that the primary feature defining the variation in the values of ΔTMLIP2 may be the greater degree of hydration in the interfacial region rather than specific interactions with the headgroups.

Correlation of PSHR and Chothia's Scales with Accessible Surface Area (ASA)

A linear relationship between the accessible surface area of amino acids and their free energy of transfer
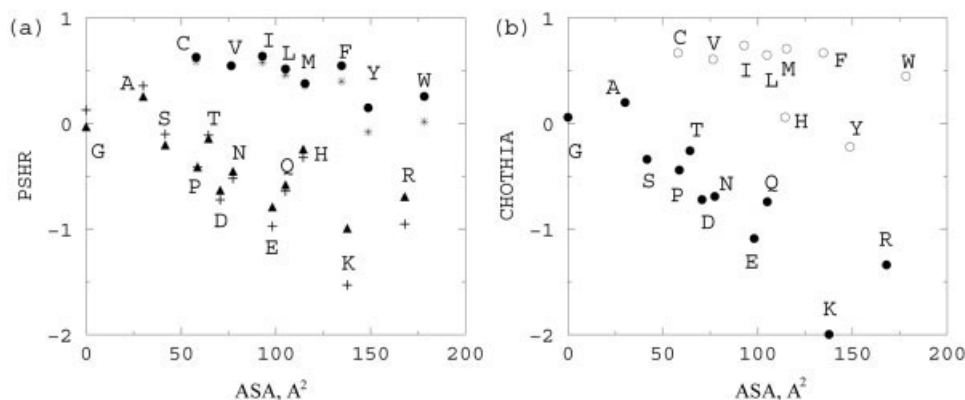
Fig. 6. Correlation of transfer free energy (kcal/mol) of amino acid residue types and accessible surface area (in Å²) in: (**a**) soluble proteins, PSHR scale (Probe radius 1.4 Å: ●, hydrophobic residues; ▲, hydrophilic and small residues. Probe radius 1.9 Å: *, hydrophobic residues; +, hydrophilic and small residues) (**b**) soluble proteins, Chothia's[30] scale.

measured experimentally from water to organic solvents was observed.[30,48,49] Comparison of free-energy scales derived from the protein structures with free-energy scales obtained from transfer experiments between water and nonpolar organic solvents showed good correlation for hydrophilic residues except Tyr and Pro, which were found more often on the surfaces of the proteins that it was expected.[30] The correlation was poor for hydrophobic residues Cys, Val, Ile, Leu, Phe, Met, and Trp, that have nearly the same energy transfer values.[30] Figure 6(a) plots PSHR transfer energies obtained from soluble protein structures with 1.4-Å and 1.9-Å probes versus solvent accessible surface area (as determined for extended chains of Gly-X-Gly peptides). This can be compared with the plot of Chothia's scale versus solvent accessible surface area [Fig. 6(b)]. Both plots show essentially the same pattern.

The linear correlation between the measured transfer free energy and the solvent accessible surface area of amino acid residues is a classical result in biophyiscs.[48,49] However, this strong correlation does not exist if the hydrophobicity scale is derived from a database of protein structures, as shown in Figure 6(a,b).

Possible reasons for this discrepancy between transfer free energies obtained from protein structures and that obtained from partition experiments have been discussed by Rose et al.[50] Additionally, proteins carry out biological functions in the living organisms. Although soluble proteins are often referred to as "globular," their surfaces are far from even. Binkowski et al.[51] showed that there are 910,379 pockets and voids in 12,177 protein structures from PDB, with approximately 15 pockets or voids for every 100 residues.[52] Functional sites are often located in these pockets and voids. Compared to the full length primary sequences of proteins, the amino acid residues forming pockets and voids are compositionally different. In particular, the aromatic residues (Phe, Trp, and Tyr) are preferentially located in pockets and voids. Interestingly, when the composition of buried and exposed residues in soluble proteins was compared, the composition of pocket residues was found to be more similar to the composition of buried, rather than exposed residues.[51] The enrichment of

functional surface pockets with hydrophobic residues and the resulting thermodynamic changes contribute to protein–ligand interactions. The deviation of PSHR and Chothia's scales from perfect linear correlation observed when using an experimentally measured scale may be a reflection of functional restraints on their sequences.

Comparison with κPROT Scale

κPROT is a lipid propensity scale derived from a sequence database analysis by Pilpel et al.[19] It is based on the idea that a higher abundance of a residue type in the TM segments of multispan proteins compared to single span proteins would indicate a propensity for this residue type to face the interior of protein and away from lipids, while a higher abundance of a residue type in the TM segments of single-span proteins indicates a higher propensity for this residue type to be exposed to the lipid phase. Presently, κPROT is the only lipophilicity scale that takes into account physico-chemical properties of the membrane with residue lipid propensities for headgroup and hydrocarbon core regions.

We found rather weak correlations between TMLIP1 and κPROT scales in both headgroup and hydrocarbon core regions with correlation coefficients of 0.57 and 0.71, respectively. The most pronounced difference between TMLIP1 and κPROT is found for aromatic residues. Phe is well sampled in TMLIP1 and is strongly lipophilic in both regions of TM helices (TMLIP1-H: 0.13, TMLIP1-C: 0.24). In the κPROT scale, Phe is lipophobic with propensities of −0.07 and −0.16 in headgroup and hydrocarbon core regions, respectively. TMLIP1 and κPROT scales assign opposite values in the hydrocarbon core region for other aromatic residues as well: Trp is neutral to lipophilic according to TMLIP1 (0.06) and is strongly lipophobic according to κPROT (−0.65), Tyr is strongly lipophobic in κPROT scale (−0.70) and neutral in the TMLIP1 scale (−0.01).

Comparative study of TM sequences of bitopic (single helix spanning) and multispan membrane proteins by Arkin and Brunger[53] showed that Phe and Trp have rather similar distributions throughout the length of TM

**TABLE IV. Summary of Lipophilicity Moment Calculations with TMLIP1,
TMLIP2, kPROT, Hydrophobicity Scales by Chothia,[30] White and Whimley,[3]
Eisenberg and Weiss,[44] and Random Scales**

| | | | Number count of ADs ||
	Average AD*	AD interval	Below 40°	Over 90°
TMLIP1	54.5	0.2 . . . 165.0	127 (49.6%)	52 (20.3%)
TMLIP2	56.2	1.0 . . . 172.0	116 (45.3%)	57 (22.3%)
kPROT	62.6	0.6 . . . 173.1	105 (41.0%)	67 (26.2%)
Chothia	61.1	0.6 . . . 179.6	108 (42.4%)	67 (26.2%)
White	66.5	0.5 . . . 178.6	101 (39.6%)	81 (31.8%)
Eisenberg	58.8	0.1 . . . 178.9	117 (45.9%)	61 (23.9%)
Random	90.1	54.1 . . . 131.3	0 (0.0%)	120 (46.0%)

*AD, Angular difference between solvent accessibility vector and lipophilicity moment in unit of degrees.

helix between both types of membrane proteins, but their overall content is higher in multispan proteins. This difference is especially pronounced in the distribution of Trp residues in the hydrophobic core region of helices in bitopic and multispan proteins: the frequency of Trp residues in the core region of bitopic proteins is at least two times smaller than in their multispan counterparts. This difference in content is likely to be reflected by kPROT scales and interpreted as a preference for the interior location. However, the bulky aromatic side chains are difficult to pack inside a helical bundle. TMLIP data show that there is a considerable fraction of Trp and Phe side chains that are accessible to the probe. In addition, the large hydrophobic surfaces of aromatic residues allow extensive contact interactions with side chains from the neighboring helices on the lipid-facing surfaces of multispan proteins, which may stabilize helical bundle structure. Such interactions involving Trp in the hydrocarbon core are often found between parallel helices that do not have an extensive interacting interface. For example, Trp 110 packs against Tyr 23 and Leu 27 of subunit I of cytochrome c oxidase from *T. thermophilus* (PDB: 1EHK). These residues reside on the parallel helices (α1 and α3) that interact mainly through their N-terminal ends and require larger side chains to provide van der Waals interactions sufficient to hold them together. In addition, there is an interhelical H-bond between carbonyl oxygen of Tyr 23 and Nϵ1 of Trp 110. This cytochrome c oxidase structure contains another Trp residue (Trp 157) that faces lipids in the hydrocarbon region of the TM region and provides extensive van der Waals as well as hydrogen bonding (Nϵ1 of Trp 157 and OG of Ser 197) interhelical interactions between helices α4 and α5. Trp 69 from QCR8 from cytochtome bc1 complex (PDB: 1KB9) from *S. cerevisiae* is an example of an involvement of Trp side chain into the interchain interactions between QCR8 and cytochtome b (residues Met 351, Ile 354, and Ile 358) in the hydrocarbon core of TM helix.

Small amino acid residues are often found at helix–helix interfaces in bitopic and multispan proteins.[54–57] The comparable distribution of Gly in both bitopic and multispan membrane proteins suggested that it is rather neutral toward facing phospholipids, resulting in kPROT propensity of −0.05 in the hydrocarbon core region. How-

ever, Gly is often buried in oligomerization interfaces of bitopic proteins such as in the case of glycophorin A, which would be scored as monomeric helices in the kPROT scale.[58] The TMLIP1 scale points to a lipophobic character of Gly residues with lipid propensity of −0.48.

Asp and Asn are highly lipophobic in the terminal kPROT scale with lipid propensity values of −0.38 and −0.73, respectively. TMLIP1-H scale indicates that these amino acid residues may be lipophilic rather than lipophobic, although the sampling is not sufficient to establish a statistically significant preference for either region. However, high-resolution X-ray structures often reveal direct polar–polar interactions between side chains of Asp or Asn and polar groups of phospholipids in the headgroup region. For example, several such interactions can be found in the structure of cytochrome bc1 complex (1KB9): Asn 7 and Asn 27 from cytochrome b interact with phosphatidylethanolamine and cardiolipin, respectively. In addition, Asn 74 forms two H-bonds with the headgroup of phosphoinositol. Recently released structure of mitochondrial ADP/ATP carrier (1OKC) also contains several examples of interactions of Asn with cardiolipin.[59] These observations suggest that Asn and Asp are likely to be lipophilic in the headgroup region. The reason for the discrepancy between TMLIP1 and kPROT may be due to the fact that these residues are located at the boundaries of the TM helices, whose locations are difficult to delineate precisely.

Prediction of TM Helix Orientations

We tested the performance of TMLIP1, TMLIP2, kPROT and three hydrophobicity scales in predicting the orientation of 256 lipid-facing TM helices. Angular differences between solvent accessibility moments and lipophilicity or hydrophobicity moments were caculated for each scale. For comparison, we generated a random scale with potentials restricted to the intervals corresponding to the highest and the lowest values of the respective headgroup or hydrocarbon core TMLIP1 potential. The random angular difference is then averaged over ten calculations for each TM helix.

We find that TMLIP1 has the best performance, with almost 50% of predicted lipophilicity moments being within 40° of the accessibility moments (Table IV). TMLIP2 lipophilicity and Eisenberg-Weiss hydrophobicity scales
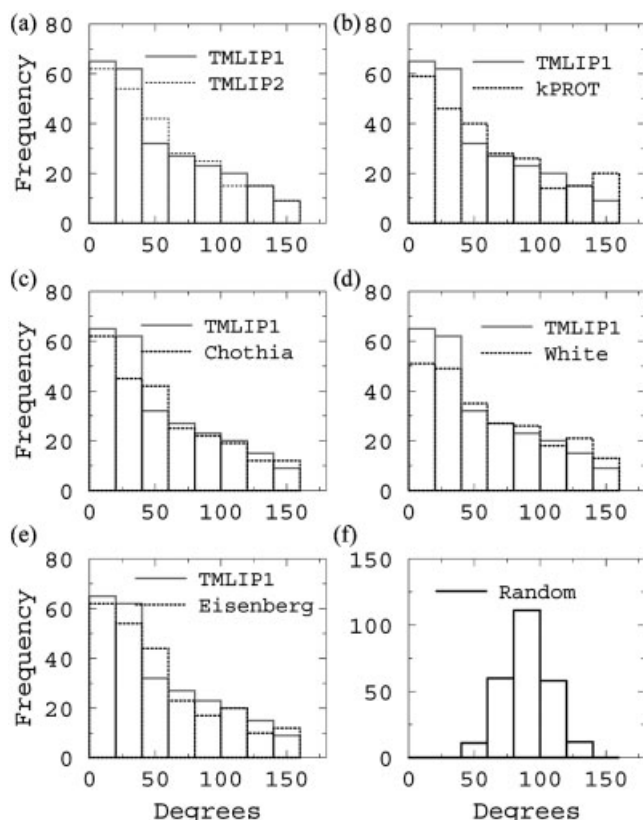
Fig. 7. Histograms of angular differences between solvent accessibility moment and lipophilicity or hydrophobicity moments. Angular differences are calculated for 256 transmembrane helices with TMLIP1, TMLIP2, kPROT,[19] Chothia,[30] White-Wimley,[3] Eisenberg-Weiss,[44] and random scales. Random angular difference is averaged over 10 calculations for each helix. TMLIP1, TMLIP2, and kPROT calculations employ combination of scales for residues in the headgroup region and the residues in the hydrocarbon core of transmembrane helix, calculations with hydrophobicity scales are run only for the hydrocarbon core region.

(b)
I.   IEV**KQ**KM**AMF**AGGASGVL**C**LIGGV**LLL**KR
II. IEVKQ**KM**AMFAGGASGVL**C**LIGGV**LLL**KR
III. IEVKQ**KM**AMF**AGGA**SGVL**C**LIGGV**LLL**KR
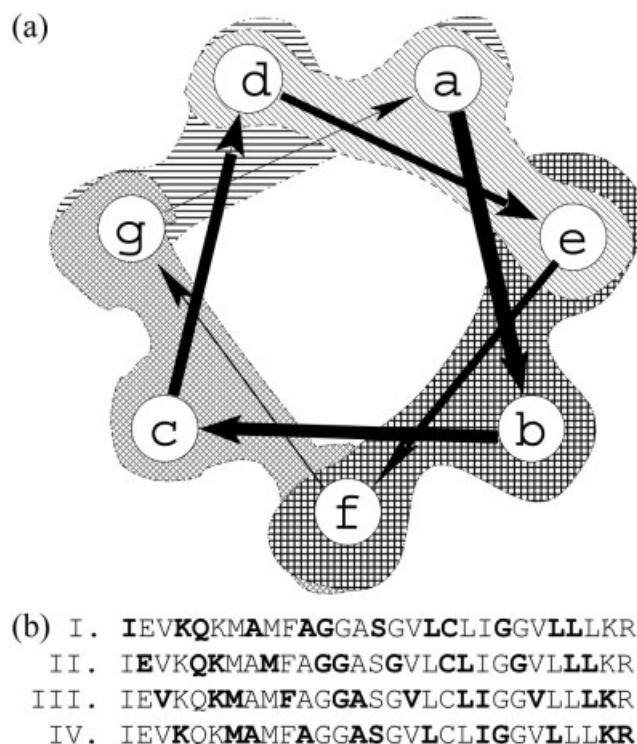IV. IEVKQ**KM**AMF**AGGASG**VL**C**LIGGV**LLL**KR

Fig. 8. **a**: Helical wheel representation of the heptad repeat. Helical positions containing residues belonging to the same helical surface are enclosed within the region of the same shading. Surface I contains residues at positions *a-d-e*, surface II: *b-e-f*, surface III: *c-f-g*, and surface IV: *d-g-a*. **b**: Sequence of helix TM3 (residues Ile 83–Phe 114, chain C) from nitrate reductase (1Q16). Highlighted residues compose helical surfaces I–IV.

predict with similar accuracy ($\sim$ 45% of predicted lipophilicity or hydrophobicity moments are within 40° from the accessibility moment). Both TMLIP scales have more predictions with small angular differences ($< 40°$) than the kPROT [Fig. 7(a, b)]. The average angular difference for 256 TM helices is 54.5° and 62.8° for TMLIP and kPROT, respectively (Table IV), while the average angular difference in calculations with randomly generated scales is around 90° (Table IV). Figure 7 shows histograms that graphically summarize the distribution of angular differences calculated with TMLIP1 and TMLIP2, kPROT, lipophilicity scales, and a random scale. The angular differences obtained with the scales [Fig. 7(a–e)] are skewed to the right with a high frequency of data in the first two bins, corresponding to angular differences of up to 40°. Angular differences obtained with a random scale are symmetrically distributed, with a maximum in the two central bins corresponding to the interval of 77°–94° [Fig. 7(f)].

Similar approaches for predicting helix orientation were previously applied to benchmark the performance of kPROT scale,[19] to predict the orientation of helices in

bacteriorhodopsin,[28] and to compare surface hydrophobicities of membrane and soluble proteins.[17] Although a clear correlation between hydrophobicity and solvent accessibility was found for soluble proteins,[17] the results for membrane proteins had lacked the same degree of clarity. Our calculation of lipid propensities shows that there are clear preferences for some amino acid residues to be on the lipid-facing surfaces of the proteins, although these preferences are not strong enough to allow unambiguous identification of all the lipid-exposed surfaces. A similar conclusion was reached by Pilpel et al.[19]

In hydrophobicity moment calculations, the hydrophobicity value of every amino acid is summed up throughout the helix. We developed another approach that is based on a coiled-coil model of interactions of α-helices in membrane proteins. Langosch and Heringa[60] examined the meshworks of the residue–residue contacts of the helix–helix interfaces of several helix-bundle proteins and found that the meshing residues at helix–helix interfaces often fit a heptad repeat pattern, with the meshing residues occurring at positions *a*, *d*, *e*, or *g*. For convenience, we use the approximation that all interacting helices assume a coiled-coil conformation and follow the heptad repeat pattern. We divided each helix into four overlapping twisted helical surfaces containing residues at the following helical wheel positions: *a-d-e*, *b-e-f*, *c-f-g* and *d-a-g*, as shown in Figure 8(a). As an example, the residues composing surfaces I to

TABLE V. Prediction of Lipid-Facing Surfaces of TM Helices in Multispan Proteins Using TMLIP1 and kPROT Scale[†]

| PDB | Number of correctly predicted helices | | N |
	TMLIP1	kPROT	
1C3W	4	3	7
1EUL	5	4	9
1FX8	1	2	6
1IWG	5	4	12
1J4N	3	4	6
1KPL	8	9	12
1KQF	3	2	5
1L9H	4	1	7
1M3X	7	6	11
1NEK	4	4	6
1OCR*	8	3	18
1PV6	4	5	12
1PW4	9	4	12
1Q16	4	3	5
Total	69	54	128
	54%	42%	

*Only subunits I, II, and III of cytochrome c oxidase were used for this study.

[†]*N*: Number of lipid-facing helices in the protein.

IV of helix TM3 of nitrate reductase (1Q16) are shown in bold in Figure 8(b). The average lipid propensity for each face is calculated and the result is compared with the percentage of probe-accessible residues on that helical face. The prediction is correct when the helical face with the highest average lipid propensity is the one with the highest surface-accessible area. We predict lipid-facing surfaces for 128 lipid-facing helices from 14 proteins (Table V, column 1). To assess the prediction objectively, the TMLIP scale is recalculated after removing each of the 27 proteins in turn, which is then used as the test example.

We then calculated average lipophilicity values for each helical surface of 128 TM helices from fourteen proteins in their monomeric form (Table V). The surfaces with larger values of average lipophilicity should have a higher probability to face phospholipids. The overall success rate of predicting the correct lipid-exposed surface is 54% for TMLIP1 scale versus 42% for kPROT scale (Table V). The TMLIP1 propensity scale was derived from the oligomeric structures of the membrane proteins when available. Exclusion of helices at protein–protein interfaces from the set of 128 helices decreased the data set to 104 helices. The performance of kPROT on this data set fell to 38%, while the performance of residue-based TMLIP scale slightly increased to 56%. Regardless of the scale used, these results are significantly better than random success rate, which is 25% (one out of four possible helical faces). The "helical face" approach allows an assessment of the collective lipophilicity of the residues that are located on the predefined helical surface, which results in a better performance.

Conclusions

We have shown that the amino acid-specific lipid interaction propensities differ strongly for different regions of the bilayer. Amino acid residues Lys, Arg, Trp, Phe, and Leu are often found at the headgroup regions, where they have a high propensity to face phospholipid headgroups and glycerol backbones. The hydrocarbon core region is enriched with hydrophobic and aromatic residues such as Ile, Leu, Val, and Phe. Small and polar amino acid residues are usually buried inside helical bundles and are strongly lipophobic. We find a good agreement between TMLIP2-C scale and hydrophobicity scales of Chothia, Eisenberg-Weiss, and White-Wimley. Thus, the interior-facing residues of membrane proteins are significantly more polar than the exterior residues. Our results indicate that the lipophobic effect may play a significant role in the folding and stability of membrane proteins. We also show that the TMLIP1 scale demonstrates the best performance in predicting lipid-facing surfaces of the TM helices in comparison with kPROT and hydrophobicity.

Note Added in Proof

While this article was in review, a related article was published by Thornton and coworkers,[61] which comes to a somewhat different conclusions concerning the distribution of polar residues in the acyl region of the bilayer. We attribute the differences in conclusion to differences in the definition of acyl versus headgroup regions in the two papers. The acyl region defined in the Thornton manuscript is, on average, broader than the core region defined in this paper, which leads to the inclusion of additional polar residues in the acyl region in their study. Otherwise, our findings are in good agreement.

REFERENCES

1. Popot JL, Engelman DM. Membrane protein folding and oligomerization: the two-state model. Biochemistry 1990;29:4031–4037.
2. Popot JL, Engelman DM. Helical membrane protein folding, stability, and evolution. Annu Rev Biochem 2000;69:881–922.
3. White SH, Wimley WC. Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 1999;28:319–365.
4. Spencer RH, Rees DC. The alpha-helix and the organization and gating of channels. Annu Rev Biomol Struct 2002;31:207–233.
5. White SH, Wimley WC. Hydrophobic interactions of peptides with membrane interfaces. Biochim Biophys Acta 1998;1376:339–352.
6. Wimley WC, Creamer TP, White SH. Solvation energies of amino acid sidechains and backbone in a family of host-guest pentapeptides. Biochemistry 1996;35:109–124.
7. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat Struct Biol 1996;3:842–848.
8. Jayasinghe S, Hristova K, White SH. Energetics, stability and prediction of transmembrane helices. J Mol Biol 2001;312:927–934.
9. Kyte J, Doolittle RF. A simple method for displaying the hydrophathic character of a protein. J Mol Biol 1982;157:105–132.
10. Eisenberg D, Weiss RM, Terwilliger TC. The helical hydrophobic

moment: a measure of the amphiphilicity of a helix. Nature 1982;299:371–374.

11. Engelman DM, Steitz TA, Goldman A. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem 1986;15:321–353.

12. von Heijne G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol 1992;225:487–494.

13. Rost B, Casadio R, Fariselli P, Sander C. Transmembrane helices predicted at 95% accuracy. Protein Sci 1995;4:521–533.

14. Rost B, Fariselli P, Casadio R. Topology prediction for helical transmembrane proteins at 86% accuracy. Protein Sci 1996;5:1704–1718.

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

16. Rees DC, DeAntonio L, Eisenberg D. Hydrophobic organization of membrane proteins. Science 1989;245:510–513.

17. Stevens TJ, Arkin IT. Are membrane proteins "inside-out" proteins? Proteins 1999;36:135–143.

18. Samatey FA, Xu C, Popot JL. On the distribution of amino acid residues in transmembrane α-helix bundles. Proc Natl Acad Sci USA 1995;92:4577–4581.

19. Pilpel Y, Ben-Tal N, Lancet D. kPROT: a knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein prediction. J Mol Biol 1999;294:921–935.

20. Beuming T, Weinstein H. A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. Bioinformatics 2004;20:1822–1835.

21. Engelman DM, Zaccai G. Bacteriorhodopsin is an inside-out protein. Proc Natl Acad Sci USA 1980;77:5894–5898.

22. Marsh D, Horvath LI. Structure, dynamics and composition of the lipid-protein interface. Perspectives from spin-labeling. Biochim Biophys Acta 1998;1376:267–296.

23. Tsai J, Taylor R, Chothia C, Gerstein M. The packing density in proteins: standard radii and volumes. J Mol Biol 1999;290:253–266.

24. Edelsbrunner H, Mucke EP. Tree-dimensional alpha-shapes. ACM Trans Graph 1994;13:43–72.

25. Edelsbrunner H, Shah NR. Incremental topological flipping works for regular triangulations. Algorithmica 1996;15:223–241.

26. Facello MA. Implementation of a randomized algorithm for Delaunay and regular triangulation in three dimensions. Comput Aided Geom D 1995;12:349–370.

27. Adamian L, Jackups JR, Binkowski TA, Liang J. Higher-order interhelical spatial interactions in membrane proteins. J Mol Biol 2003;327:251–272.

28. Donnelly D, Overington JP, Ruffle SV, Nugent JH, Blundell TL. Modeling α-helical domains: the calculations and use of substitution tables for lipid-facing residues. Protein Sci 1993;2:55–70.

29. Komiya H, Yeates TO, Rees DC, Allen JP, Feher G. Structure of the reaction center from *R. sphaeroides* R-26 and 2.4.1: symmetry relations and sequence comparisons between different species. Proc Natl Acad Sci USA 1988;85:9012–9016.

30. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol 1987;196:641–656.

31. Planqe MRR, Killian JA. Protein-lipid interactions studied with designed transmembrane peptides: role of hydrophobic matching and interfacial anchoring (Review). Mol Membr Biol 2003;20:271–284.

32. Yau W-M, Wimley WC, Gawrisch K, White SH. The preference of tryprophan for membrane interfaces. Biochemistry 1998;37:14713–14718.

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

34. Liu W, Eilers M, Patel AB, Smith SO. Helix packing moments reveal diversity and conservation in membrane protein structure. J Mol Biol 2004;337:713–729.

35. Choma C, Gratkowski H, Lear JD, DeGrado WF. Asparagine-mediated self-association of a model transmembrane helix. Nat Struct Biol 2000;7:161–166.

36. Gratkowski J, Lear JD, DeGrado WF. Polar side chains drive the association of model transmembrane peptides. Proc Natl Acad Sci USA 2001;98:880–885.

37. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM. Interhelical hydrogen bonding drives strong interactions in membrane proteins. Nat Struct Biol 2000;7:154–160.

38. Adamian L, Liang J. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. Proteins 2002;47:209–218.

39. Lear JD, Gratkowski H, Adamian L, Liang J, DeGrado WF. Position-dependence of stabilizing polar interactions of asparagine in transmembrane helical bundles. Biochemistry 2003;42:6400–6407.

40. Ruan WM, Lindner E, Langosch D. The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis. Protein Sci 2004;13:555–559.

41. Ruan WM, Becker V, Klingmuller U, Langosch D. The interface between self-assembling erythropoietin receptor transmembrane segments corresponds to a membrane-spanning leucine zipper. J Biol Chem 2004;279:3273–3279.

42. Stevens TJ, Arkin IT. Substitution rates in α-helical transmembrane proteins. Protein Sci 2001;10:2507–2517.

43. Donnelly D, Johnson MS, Blundell TL, Sounders J. An analysis of the periodicity of conserved residues in sequences alignments of G-protein coupled receptors. FEBS Lett 1989;251:109–116.

44. Eisenberg D, Weiss RM, Terwilliger TC, Wilcox W. Hydrophobic moments and protein structure. Faraday Symp Chem Soc 1982;17:109–120.

45. Hobohm U, Sander C. Enlarged representative set of protein structures. Protein Sci 1994;3:522–524.

46. Rees DC, Chirino AJ, Kim K-H, Komiya H. Membrane protein structure and stability: implications of the first crystallographic analyses. In: White SH, editor. Membrane protein structure. New York, Oxford: Oxford University Press; 1994. p 3–26.

47. Braun P, von Heijne G. The aromatic residues Trp and Phe have different effects on the positioning of a transmembrane helix in the microsomal membrane. Biochemistry 1999;38:9778–9782.

48. Chothia C. Hydrophobic bonding and accessible surface area in proteins. Nature 1974;248:338–339.

49. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.

50. Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. Science 1985;229:834–838.

51. Binkowski TA, Adamian L, Liang J. Inferring functional relationships of proteins from local sequence and spatial surface patterns. J Mol Biol 2003;332:505–526.

52. Liang J, Dill K. Are proteins well-packed? Biophys J 2001;81:751–766.

53. Arkin IT, Brunger AT. Statistical analysis of predicted transmembrane α-helices. Biochim Biophys Acta 1998;1429:113–128.

54. Eilers M, Patel AB, Liu W, Smith SO. Comparison of helix interaction in membrane and soluble α-bundle proteins. Biophys J 2002;82:2720–2736.

55. Dawson JP, Weiner JS, Engelman DM. Motifs of serine and threonine can drive association of transmembrane helices. J Mol Biol 2002;316:799–805.

56. Russ WP, Engelman DM. The GxxxG motif: a framework for transmembrane helix-helix association. J Mol Biol 2000;296:911–919.

57. Senes A, Gerstein M, Engelman DM. Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. J Mol Biol 2000;296:921–936.

58. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science 1997;276:131–133.

59. Pebay-Payroula E, Dahout-Gonzalez C, Kahn R, Trezeguet V, Lauquin GJ, Brandolin G. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature 2003;426:39–44.

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

61. Eyre TA, Partridge L, Thornton JM. Computational analysis of α-helical membrane protein structure: implications for the prediction of 3D structural models. Protein Eng Des Sel 2004;17:613–624.