PROTEINS: Structure, Function, and Genetics 36:135–143 (1999)

Are Membrane Proteins "Inside-Out" Proteins?

Timothy J. Stevens and Isaiah T. Arkin*
Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, United Kingdom

ABSTRACT **One of the central paradigms of structural biology is that membrane proteins are "inside-out" proteins, in that they have a core of polar residues surrounded by apolar residues. This is the reverse of the characteristics found in water-soluble proteins. We have decided to test this paradigm, now that sufficient numbers of transmembrane α-helical structures are accessible to statistical analysis. We have analyzed the correlation between accessibility and hydrophobicity of both individual residues and complete helices. Our analyses reveal that hydrophobicity of residues in a transmembrane helical bundle does not correlate with any preferred location and that the hydrophilic vector of a helix is a poor indicator of the solvent exposed face of a helix. Neither polar nor hydrophobic residues show any bias for the exterior or the interior of a transmembrane domain. As a control, analysis of water-soluble helical bundles performed in a similar manner has yielded clear correlations between hydrophobicity and accessibility. We therefore conclude that, based on the data set used, membrane proteins as "inside-out" proteins is an unfounded notion, suggesting that packing of α-helices in membranes is better understood by maximization of van der Waal's forces, rather than by a general segregation of hydrophobicities driven by lipid exclusion. Proteins 1999;36:135–143.** © 1999 Wiley-Liss, Inc.

Key words: protein structure; protein folding; hydrophobicity; α-helix

INTRODUCTION

With the availability of whole organism genome sequences, it has become clear that membrane proteins are extremely abundant (20–50% of known proteins).[1] Their importance in biomedicine is perhaps even more so, as the majority of pharmaceuticals in use today target membrane proteins. However, despite their medical significance and natural abundance, there are less than a dozen listed high resolution structures of integral membrane proteins in the Brookhaven Protein Data Bank (PDB). This is a result of the fact that structures of transmembrane domains of membrane proteins are notoriously difficult to determine experimentally.

As the application of experimental methods for elucidating membrane protein structures has encountered difficulties, the research community has directed its efforts at analyzing membrane proteins on a theoretical basis (model building), aided by low resolution structural data (e.g., mutagenesis data). One of the major paradigms underlying such efforts is that membrane proteins, due to the restrictions enforced by the lipid bilayer, are "inside-out" proteins: i.e., proteins that contain a polar core and an apolar exterior.[2,3,4] This situation is the opposite from that found in soluble proteins, in which hydrophobic collapse of the core is thought to drive protein folding, resulting in a protein with an apolar core and a polar exterior.[5] The inside-out paradigm has guided many model building efforts, by way of positioning particular polar amino acids in the core of the transmembrane bundle. A rigorous mathematical manifestation of this paradigm has resulted in the hydrophobic moment concept.[6]

Structures of transmembrane domains can be subdivided into three categories: those with a β-barrel structure, as is found in the porin class of proteins, those which traverse the lipid bilayer with a single α-helix (bitopic), and those that cross the bilayer with multiple α-helices (polytopic), i.e., an α-helical bundle. Transmembrane α-helical domains can be predicted by hydropathy analysis, suggesting that membrane domain assembly may follow a different set of rules to soluble proteins. We have decided to examine whether the database of α-helical membrane proteins structure (containing a total of 61 α-helices) obeys the inside-out paradigm. To that end, using the current albeit limited data set, we have shown that the rules governing the direction of amphipathicity (as described by the hydrophilic vector) of a helix, or the solvent accessibility of particular residues, are inconsistent with an inside-out morphology. The data suggests that within the hydrophobic environment of the lipid bilayer, the α-helical orientation is governed more by factors such as maximization of van der Waal's forces, than by a general segregation of hydrophilicities driven by lipid exclusion.

MATERIALS AND METHODS

Protein Database Selection

The subset of the Brookhaven Protein Data Bank (PDB) chosen for this study are those structures of transmembrane proteins that are anchored in a lipid bilayer by a polytopic α-helical bundle. The currently known PDB

Abbreviations: TM, transmembrane; PDB, protein data bank; TMαH, transmembrane α-helix

Grant sponsor: The Wellcome trust; Grant sponsor: The Biotechnology and Biological Sciences Research Council.
*Correspondence to: Isaiah T. Arkin, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, UK. E-mail: sa232@cam.ac.uk
Received 16 November 1998; Accepted 12 March 1999

structures which lie within this category are: *Bos taurus* Cytochrome c Oxidase (1occ),[7] *Paracoccus denitrificans* Cytochrome c Oxidase (1ar1),[8] *Gallus gallus* Cytochrome bc_1 complex (1bcc),[9] *Halobacterium halobium* Bacteriorhodopsin (2brd),[10] *Rhodopseudomonas acidophia* Light Harvesting Complex (1kzu),[11] *Homo sapiens* lycophorin A (1afo),[12] *Rhodopseudomonas viridis* Photosynthetic Reaction center (1prc),[13] *Rhodopseudomonas spheroids* Photosynthetic Reaction center (1pcr),[14] and *Streptomyces lividans* Potassium channel.[15] The Brookhaven PDB was searched for any recent accessions which would fit the above criteria by using both the structural classification of proteins database (SCOP)[16] and keyword text searches against the whole PDB. There are seven non-homologous structures which contain 61 distinct transmembrane α-helices. Where there was a choice between two or more homologous structures, the structure with the largest number of TM α-helices (e.g., when choosing between bovine and bacterial cytochrome *c* oxidase) or with the best resolution has been chosen (e.g., for the two photosynthetic reaction centers).

For purposes of comparison, the structures of soluble helical bundles were obtained, as with the membrane proteins, with the aid of the SCOP database. The structural classification used was the all α bundle. Again, as with the membrane proteins, a data set of non-homologous proteins was used. These are: *Locusta migratoria* Apolipophorin III (1aep),[17] *Ectatomma tuberculatum* Ectatomin (Ant venom) (1eci),[18] *Homo sapiens* Ferritin (1fha),[19] *Escherichia coli* ROP (1rop),[20] *Homo sapiens* Apolipoprotein E3 (1lpe),[21] and *Escherichia coli* Cytochrome *b*562 (256b).[22]

Delineation of Transmembrane α-Helices

PDB structure files were created for each TMαH (transmembrane α-helix) domain. The PDB structure for each complete protein structure was edited to omit any extramembranous elements, forming the database of residues to which further analysis is to be applied. The helical regions were determined by the secondary structure assignment of the program DSSP.[23] A specially written program dismisses any hydrophobic helical regions that are too short to fit in a lipid bilayer (12 residues).[24] Small non-helical kinks within the TM domain were accounted for by allowing the presence of one or two linker residues between helical regions. The output TMαH files were inspected visually using 3D viewing software to check that no obvious TMαHs had been incorrectly dismissed. This check also illustrates when automation has left long aqueous α-helices in the structure files. These are then removed by manual editing.

Each TM α-helix was determined on the basis of hydropathy and helical length. Hydropathy analysis (GES scale[25]) was performed on each TM α-helix to define the ends of the hydrophobic helical region. TMαH ends are characterized by a marked change in hydrophobicity, often with the presence of a charged residue which may interact with the lipid head-groups. Only if the helix is of sufficient size to fit in a lipid bilayer were its residues considered for further analysis. Visual inspection of the selected helices shows that only helices which would be expected to lie within the lipid bilayer were selected by this automated process.

The final delineated helices are listed below for the transmembrane database:

1occ: A15–A37, A56–A83, A102–A114, A145–A165, A184–A210, A234–A254, A271–A283, A303–A325, A337–A357, A380–A397, A414–A426, A453–A473, B28–B46, B64–B78, C21–C35, C42–C52, C83–C101, C136–C147, C159–C176, C197–C220, C244–C256, D81–D98, G18–G33, I21–I53, K16–K33, L21–L40 & M16–M34;
1bcc: C34-C54, C88–C97, C114–C130, C179–C201, C230–C246, C289–C301, C325–C339, C348–C373, D204–D222, E38–E51, G48–G61 & J20–J31;
1prc: L34–L49, L89–L102, L117–L134, L179–L189, L232–L249, M56–M75, M112–M129, M144–M161, M202–M216, M266–M284 & H17–H31;
2brd: 11–29, 42–63, 83–101, 108–127, 135–156, 177–191 & 206–224;
1bl8: A28–A50 & A90–A113;
1kzu: A13–A35 & B21–B34;
1afo: 73–95.

Helices from the soluble helices database are listed below:

1eci: A6–A35 & B5–B17;
1lpe: 26–41, 46–78, 88–123 & 132–162;
1rop: 5–28 & 33–55; 8–25; 37–63, 70–83, 95–121 & 132–154;
1fha: 15–41, 50–76, 97–123 & 128–158;
256b: 4–19, 24–40, 57–80 & 85–105.

Residues in close proximity to cofactor or prosthetic group atoms were excluded from the set of TM residues. Their interactions do not represent helix-helix or helix-lipid interactions, being specific to the biological function of the protein. The presence of any atom of a residue side chain within 3 Å of a hetero-atom (hydrogen atoms are implicit) led to the deletion of the whole residue.

Accessibility Assignments

The solvent accessibility of a particular residue is a measure of the amount of van der Waal's surface area that can be accessed by a sphere of specified diameter. For this study, the values of accessibility are those defined by the output of the program MSRoll.[26] When determining the accessibility of residues within TM domains, the PDB file for the whole protein complex, including prosthetic groups, was taken into account. However, peripheral molecules required for crystallization, e.g., lipids and sulfate anions, were excluded. Consequently, the effects of aqueous domain and non-helical residues were taken into account, even though the only directly studied residues were those within the TMαHs.

The accessibility range is inherently larger for residues with a larger van der Waal's surface area. Therefore, when comparing different amino-acid residues it is necessary to express residue exposure as the proportion of maximum

accessibility. Solvent accessibility values for the residues within the data set were grouped according to residue type, for each individual membrane bundle and the data as a whole. The range of accessibility values for each residue type were grouped into ten bins (each corresponding to a range of 0.1). The number of residues in each group was then counted and presented as a histogram. The analysis was performed on the TMαHs and on the helices of soluble bundles.

Hydrophilic and Accessibility Vectors

The hydrophilic vector of a helix is a measure of the helix's amphipathic nature—the sum of the hydrophilic vectors for each of the residues within the helix. The accessibility vector of a helix is calculated in an analogous manner to the hydrophilic vector. The accessibility vector of a given helix is the sum of the accessibility vectors of its residues. The vector represents the side of the helix which is most exposed to the membrane bilayer (solvent exposed). The hydrophilic and accessibility vectors, for a given residue, are in a direction parallel to the vector between the α-carbon and the geometric center of the side chain and have magnitudes proportional to residue hydrophobicity and accessibility respectively. Here we are using the GES scale to generate hydrophobicity values (such that the hydrophobic residues have a negative value) and the program MSRoll[26] (as described above) to calculate solvent accessibility figures. The two helix vector properties (the sums of the residue vectors) are then resolved in a plane perpendicular to the helical axis. In order to correlate the direction of the hydrophilic vectors with the accessibility vectors, we have tabulated the dot product between these vectors for each of the helices in both the TMD and aqueous bundle databases.

Residues that were removed from consideration during the initial accessibility analysis, due to proximity to cofactors, were included in the vector analysis. Visual inspection of the hydrophilic vectors with and without these residues shows that there is little difference when these residues are removed. In addition, there is no difference to the overall hydrophobic pattern. To show the overall pattern of amphipathicity for the bundle and to determine whether it has an influence upon the oligomerisation, the whole oligomeric complex of a helical bundle is required. Thus, repeated helices of oligomeric complexes have not been excluded from the vector analysis (in contrast to the residue accessibility analysis). However, the dot products between hydrophilic and accessibility vectors are calculated for only non-homologous helices.

RESULTS

Database Selection

From a total of seven structures (1occ, 1bcc, 1prc, 2brd, 1bl8, 1kzu, and 1afo), a total of 61 individual transmembrane α-helices were selected using protocols described in the METHODS section. As a representative, Fig. 1(**a**) and (**b**) presents both the starting structure and the resulting delineated structure resulting from the procedures used to separate the TM domains of the photosynthetic reaction center (1prc).

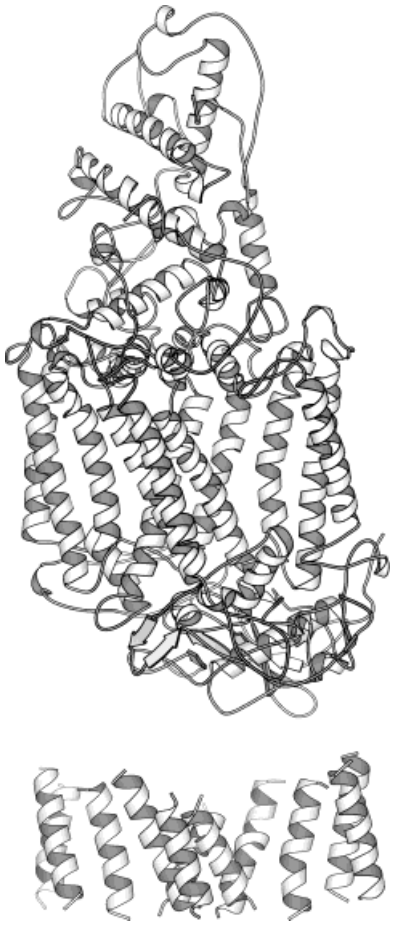


Fig. 1. Starting structure of the photosynthetic reaction center 1prc[13] (top) and the structure used in our analysis (bottom), after removal of all extra-membranous elements. The figure was generated using Molscript.[27]

Table 1 presents both the absolute and relative compositions of the amino acids in the membrane and soluble helical databases used in the analysis. The amino acid composition of the transmembrane protein database is similar to that found in much larger putative transmembrane protein sequence databases.[28] This result may indicate that, although the number of proteins in our database is relatively small, it may serve as an adequate representation to a larger subset of membrane helical bundles. Following the results depicted in Table 1, we have decided to exclude from our analysis any residue that did not appear more then six times in the transmembrane (Q, E, K, D & R) or soluble (P) protein databases.

Accessibility

Fig. 2 depicts a surface histogram plot representation of the accessibility of the particular residues found in transmembrane and soluble protein databases as a function of residue hydrophobicity. Note that the histogram groups correspond to the proportion of maximum possible accessi-

TABLE I. Counts and Relative Abundance of Amino Acid Residues in the Helices of the Transmembrane Domain (TMD) and Aqueous All-α Bundle (AAB) Databases†

Amino Acid Type	Abundance in TMD Data		Abundance in AAB Data	
	Count	%	Count	%
F	124	10.9	11	2.1
M	62	5.4	13	2.5
I	115	10.1	18	3.5
L	224	19.7	72	13.8
V	112	9.8	22	4.2
C	14	1.2	8	1.5
W	41	3.6	6	1.2
A	130	11.4	64	12.3
T	77	6.8	30	5.8
G	98	8.6	10	1.9
S	56	4.9	22	4.2
●P	23	2.0	3	0.6
Y	27	2.4	15	2.9
H	10	0.9	16	3.1
○Q	5	0.4	37	7.1
N	10	0.9	30	5.8
○E	2	0.2	48	9.2
○K	3	0.3	40	7.7
○D	5	0.4	25	4.8
○R	0	0.0	30	5.8

†Amino acids are listed in the table in order of descending hydrophobicity according to the GES scale.[29] ○ and ● indicate residues that are excluded (due to low abundance) from the analysis of TMDs and AABs respectively.

bility, such that both glycine and tryptophan have the same accessibility potential. Also, the histogram for each amino acid is normalized according to its frequency, so that residues with contrasting abundances can be compared. Upon comparing the results from Fig. 2, several major conclusions can be derived:

1. Inspection of Fig. 2**a** points to the lack of any general correlation between hydrophobicity and accessibility in the transmembrane protein database. Thus, no residue exhibited any marked preference towards the interior of the structure or towards the exterior in the transmembrane protein database.
2. Results from the soluble protein database contrast the results for transmembrane helical database (Figure 2). Hydrophobic residues were found to segregate preferentially in the interior of the structures (exhibiting low accessibility), while polar resides were found preferentially in the exterior of the complex (exhibiting high accessibilities).

Inspection of data arising from each of the transmembrane structures individually (data not shown) demonstrates that the final results are consistent with data arising from any individual structure, thus providing statistical cross validation. In other words, results are self-consistent and are not dominated by any one protein structure. This result lends more credence to our assertion that this database does represent an adequate statistical sampler of transmembrane proteins.

Hydrophilic Vector Calculation

Fig. 3 depicts the result of the hydrophilic vector analysis on the water soluble helical bundles. Fig. 4 represents the results of the same analysis for the transmembrane α-helical bundles. Visual inspection clearly points out that in water-soluble helical bundles, all of the hydrophobic vectors are pointing outwards from the helix bundle. In the transmembrane helical bundles, no regularity in the positions of the hydrophilic vectors is observed, with the possible exception of bacteriorhodopsin.

Hydrophobicity and Solvent Exposure Vector Correlation

Fig. 5 and 6 are histograms to show the dot product between the hydrophilic and solvent accessibility vectors for each helix in the aqueous bundle and TMD databases. The dot product is used to show how the vectors are correlated as it gives an indication of both the relative directions and magnitudes of the vectors. Large values for the dot product indicate that the vectors have significant magnitudes and face in the same general direction, for positive values or in opposite directions, for negative values. A dot product close to zero indicates that the vectors are either orthogonal or that at least one vector has a small magnitude.

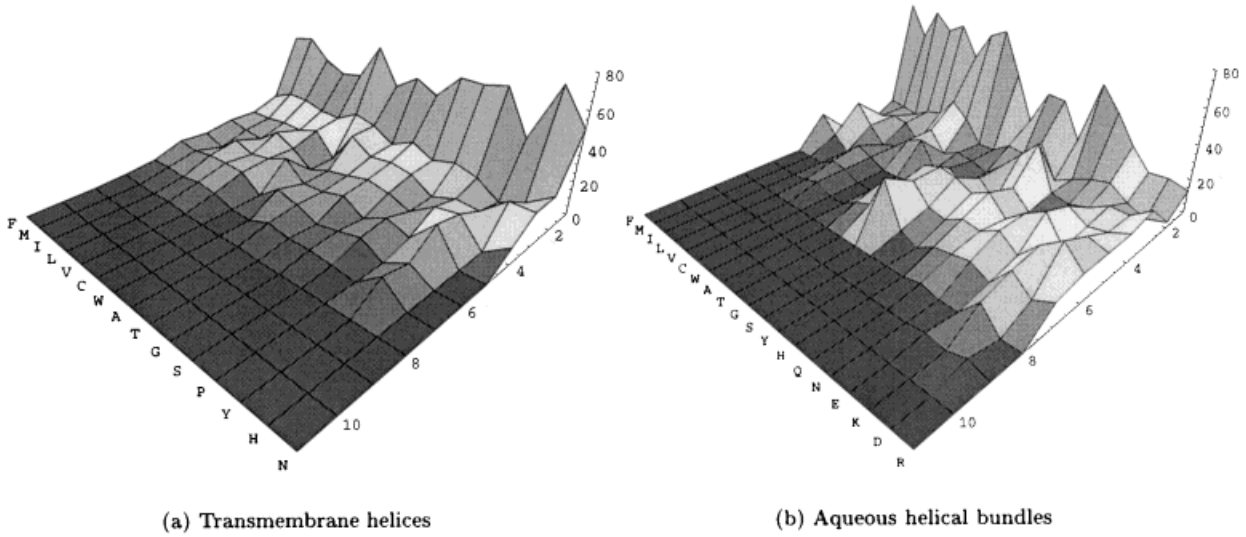
The histogram presented on Fig. 5 shows that for aqueous helical bundles, there is a very strong correlation between accessibility and hydrophilic vectors; the dot product of these vectors are generally large and always positive. For the water-soluble helical bundles, the accessibility and hydrophilic vectors point in the same direction, indicating (as would be expected) that the hydrophilic face of the helices is the solvent exposed face.

Fig. 6 shows that for TMD bundles there is a very weak correlation between vectors; the dot product between vectors are small and spread either side of zero. This shows that the solvent exposed face of the helix and the hydrophilic vector are poorly correlated. Of note is a slight bias for helices to face away from the solvent exposed surface (i.e., a negative dot product).

DISCUSSION

Parameters Governing Membrane Protein Structure

A lipid bilayer is characterized by chains of aliphatic carbon atoms, segregated from an aqueous environment by a lack of favorable solvation interactions. Transmembrane α-helices, like the lipids that surround them, are hydrophobic. It is this hydrophobicity which ensures the stability of the membrane domain; the transition from lipid to water would be energetically very costly. The hydrophobic environment of the lipid bilayer contains no groups capable of hydrogen-bonding, hence the protein must self-satisfy its backbone hydrogen bonding potential. Thus, transmembrane structures are invariably β-barrels or bundles of α-helices. Typical ΔG values for an oil to water transition for a TMαH are in the order of -40 kcal mol^{-1}.[28]



(a) Transmembrane helices (b) Aqueous helical bundles

Fig. 2. Accessibility histograms of amino acid residues, normalized for maximal possible accessibility, and abundance (see METHODS) for transmembrane helices (a) and for the soluble helical database (b). Residues are aligned according to their relative hydrophobicity based on the GES scale. The vertical scale represents the % of residues in a particular accessibility group.[25]

While maintaining their overall hydrophobicity, many TMαHs do contain polar and occasionally ionizable residues (see Table 1). The abundance of hydrophilic residues (those with a GES hydrophobicity greater than or equal to that of glycine) in the data used is about 20% of that of the hydrophobic residues. Traditionally, these hydrophilic residues are thought to be accommodated in transmembrane protein elements by sequestering them into the core of the protein.[2,3,4] This model is described as inside-out, whereby polar residues are found in the interior of the protein, and apolar residues are found in the exterior, forming favorable contacts with the lipid bilayer. This suggests that the driving force behind helix-helix association in the membrane is exclusion from the membrane lipids. This morphology is the reverse of that found in soluble protein, in which an apolar core exists, and in fact is thought to drive the folding of soluble proteins by hydrophobic collapse.[5] Thus, it is envisioned that both membrane and soluble helices contain amphipathic helices forming a bundle by sequestering either the polar or apolar face of the helix, respectively. Utilizing the concept of membrane proteins as inside-out proteins has led to model building efforts based on locating the polar residues on a helical wheel diagram and assigning it to the core of the protein.

It is interesting to note that the inside-out paradigm of membrane proteins conflicts with the two-stage model for membrane proteins' folding and oligomerization,[30,31] which states that membrane proteins fold (or oligomerize) in two stages:

- In the first stage, individually stable transmembrane α-helices form.
- In stage two, the individually stable helices associate to form the final structure.

Thus, independent stable transmembrane α-helices are at the core of the two-stage model for membrane protein folding and oligomerization. Clearly such helices would not be stable and would not form independent folding units if they carried a substantial amphipathic nature manifesting itself in a measurable net hydrophilic vector. To this end, mutations of hydrophobic → strongly hydrophilic helix residues are known to generally disrupt membrane domain formation.[32]

The inside-out paradigm of membrane proteins also conflicts with experimental data, in which mutagenesis was used to determine which residues are located in the protein-protein interface. Both in the case of the dimeric human glycophorin A[32,33] and the pentameric phospholamban,[34,35] sensitivity towards substitution was observed, not so much as a function of hydrophobicity, but rather as a function of the structure of the amino acid.[36]

Previous attempts to characterize this structural phenomenon on a single known structure (photosynthetic reaction center) has lead Rees and co-workers to conclude that in this protein a polar core does exist.[3] Our study is dramatically distinct in several critical aspects:

1. The database is about five times larger both in terms of number of helices and number of individual proteins. Thus the study is able to determine whether the hydrophobic organization of an individual structure is typical of the whole class of structures.
2. In our analysis of individual residues we have not taken into account residues which are involved in cofactor binding (which are mostly polar if not ionizable). It is our contention that such residues do not reflect protein-protein, nor protein-lipid contacts, and as such should be excluded from residue analysis.

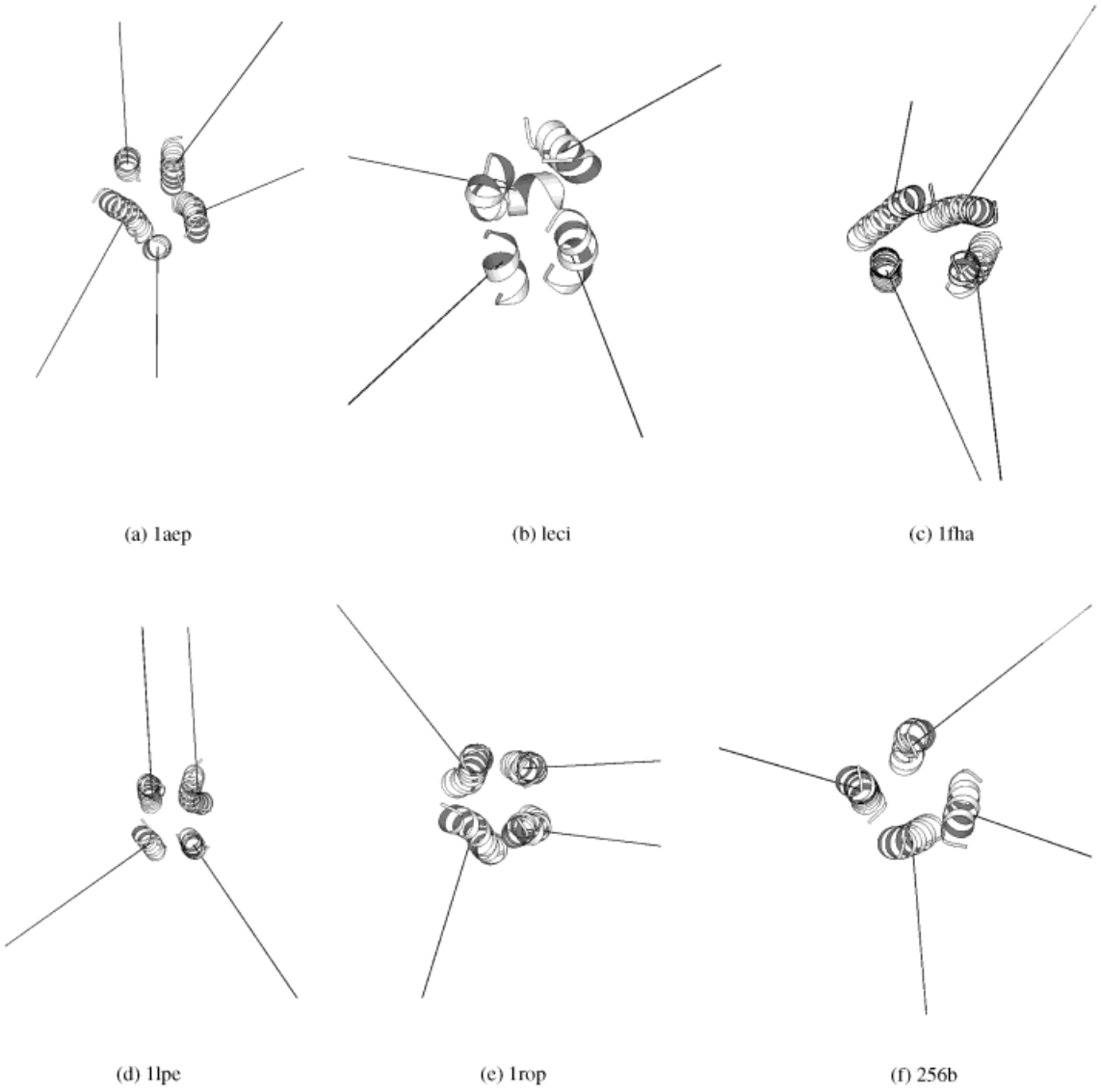


Fig. 3. Hydrophilic vector analysis for water soluble helical bundles.

3. The last distinction is that our analyses have focused on individual residues as well as complete helices, and did not involve an empirical cut-off of accessibility into three categories.

The hypothesis that membrane proteins are inside-out proteins has so far escaped any statistical validation, as the database of membrane proteins up until 2 years ago was prohibitively small. However, presented with several new structures within the past 2 years, we have set forth to determine whether the inside-out folding principles are in fact found in structures of transmembrane helical bundles. In order to address the question of whether hydrophobicity correlates with the lipid exposure of residues, we calculated the solvent accessibility (exposure) for each helix residue. We then grouped the accessibilities for the different residue types, so that the distribution of a residue can be compared with other residues.

Soluble Proteins Analysis

Initially, we examined a set of soluble helical bundles in order to ascertain the validity of our analysis. The accessibility plots for soluble helical bundles clearly show the presence of a hydrophobic core with the presence of two classes of residues (Fig. 2). Hydrophobic residues almost exclusively occur at low accessibility and hydrophilic residues that are distributed around the periphery of the helical bundles. A strong correlation between the direction of the hydrophilic and accessibility vectors (Fig. 5) of aqueous bundles provides an alternative illustration of the expected hydrophobic core of water-soluble proteins.

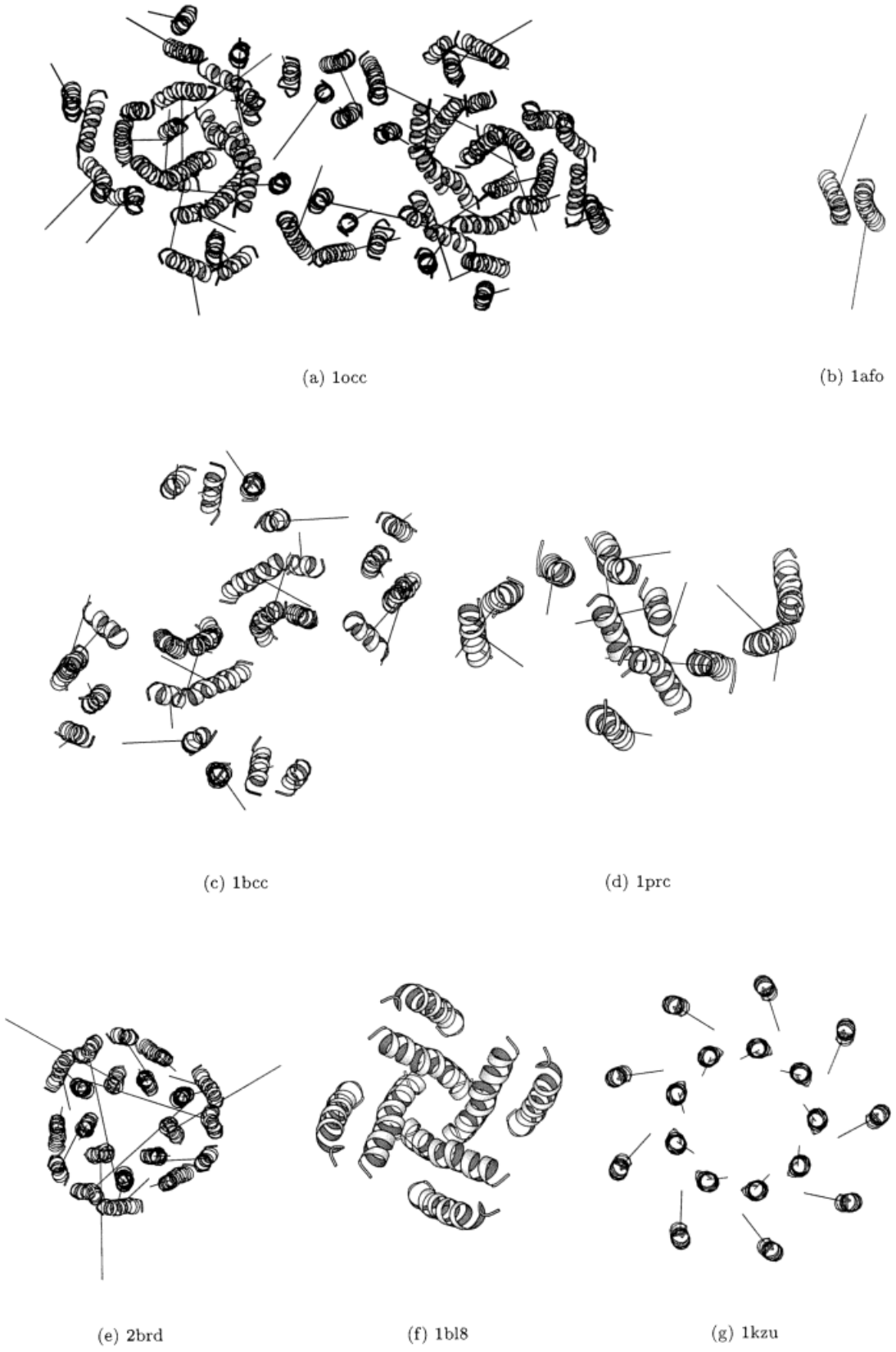


Fig. 4. Hydrophilic vector analysis for transmembrane helical bundles.

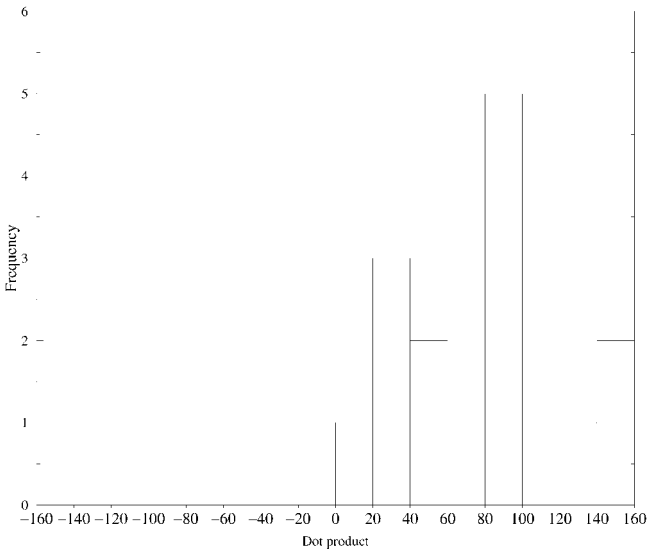


Fig. 5. A histogram showing the distribution of the dot products for accessibility and hydrophilic vectors for the helices in the aqueous bundle database. The vertical scale represents the number of helices within each histogram bin. The horizontal scale indicates the histogram bins.

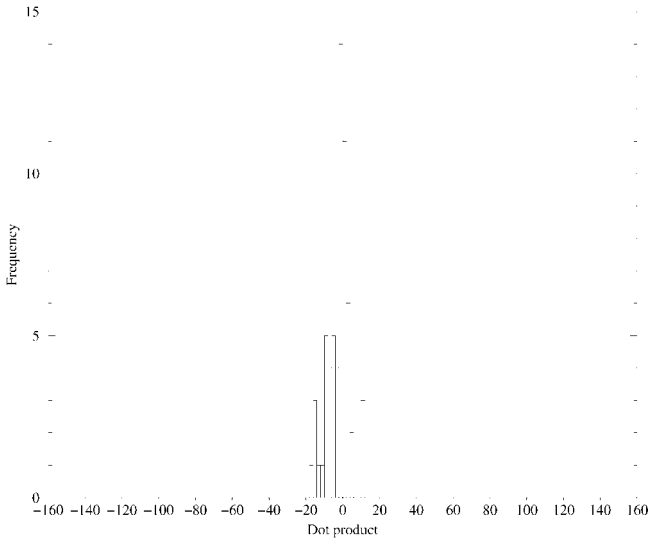


Fig. 6. A histogram showing the distribution of the dot products for accessibility and hydrophilic vectors for the helices in the TMD bundle database. The vertical scale represents the number of helices within each histogram bin. The horizontal scale indicates the histogram bins.

Fig. 2 also suggests that for the hydrophilic residues, the greater the hydrophilicity, the greater the average accessibility. The above results illustrate that the approach we have undertaken is useful, in the sense that well known phenomena are observed, and we conclude that it provides a good basis for comparison when analyzing membrane helices.

Transmembrane Proteins Analysis

The accessibility plots for the hydrophobic transmembrane α-helices show that the hydrophilic residues (Fig. 2a), in general, have no greater predisposition for regions of low solvent accessibility than the hydrophobic residues. In particular, the most abundant hydrophilic residues, the hydroxyl-bearing residues (Ser, Thr & Tyr) have distributions indistinguishable from the hydrophobic residues.

A spread of the dot products (of accessibility and hydrophilic vectors) either side of zero for TMD helices illustrates that the two vectors are poorly correlated. This indicates that the hydrophilic vector of a helix is not a sound basis for the prediction of helical twist. It would seem, from these results, that hydrophobicity is not a good indicator of residue position or helix orientation, perhaps indicating that the differences between the two hydrophobic environments of protein and lipid is insignificant compared to other helix-orienting factors. We can therefore conclude that, based on the database used in this study, membrane proteins in general do not exhibit an inside-out character.

On the other hand, we do expect to find catalytic residues, or residues which are involved in cofactor binding (such as lysine 216 in bacteriorhodopsin), which are mostly polar (if not ionizable) residues, located preferentially in the protein core. However, these residues do not represent a contribution to protein stability through protein-protein or protein-lipid interactions, but rather a functional requirement.

While the correlation between accessibility and hydrophobicity is a measure of the propensity of individual amino-acids, the vector properties are for the helix as an average. Yet both analyses lead to the same conclusion. While the hydrophilic vector of water-soluble helices are all directed outward from the helix bundle, signifying an apolar interior, the hydrophilic vectors of transmembrane helical bundles are in general randomly oriented (with the possible exception of bacteriorhodopsin and photosynthetic reaction center). This would once again indicate that hydrophobicity alone cannot be the driving force behind the packing of α-helices in lipid bilayers. Indeed, it can be imagined that solvation by lipids and the formation of protein-protein interactions is not greatly affected by the sparse polar residues in membrane helices. The important stabilizing interactions of polar residues, solvation by water, and the formation of hydrogen bonds, do not occur in the membrane domain. The fact that the hydrophilic vectors of bacteriorhodopsin are oriented inwardly may be the causative agent of its extraordinary stability, as compared to other membrane proteins.[37]

CONCLUSION

Our data shows that there is no clear correlation between hydrophobicity and a bias for the exterior surface of a helical bundle. This suggests that the driving force for a particular orientation of oligomerization is the formation of a compact helical bundle with maximized van der Waal's interactions which complies with the constraints imposed by the extra-membranous protein loops. There is no indication that membrane proteins, as a class, exhibit inside-out hydrophobic organization. Functional residues aside, it is

clear that a polar group of atoms has no preference for the hydrophobic lipid environment or the similarly hydrophobic protein environment. Thus, residues reside where they fit best to contribute to stability of the protein (or complex).

ACKNOWLEDGMENT

This work was supported by grants from the Wellcome trust and the Biotechnology and Biological Sciences Research Council to I.T.A.

REFERENCES

1. Arkin IT, Brunger AT, Engelman DM. Are there dominant membrane protein families with a given number of helices? Proteins struct. func. genet. 1997;28:465–466.
2. 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(1):125–142.
3. Rees DC, DeAntonio L, Eisenberg D. Hydrophobic organization of membrane proteins. Science 1989;245(4917):510–513.
4. Engelman DM, Zaccai G. Bacteriorhodopsin is an inside-out protein. Proc. Natl. Acad. Sci. USA 1980;77(10):5894–5898.
5. Dill KA, Chan HS. From Levinthal to pathways to funnels. Nat. Struct. Biol. 1997;4(1):10–19.
6. Eisenberg D, Weiss RM, Terwilliger TC. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. Nature 1982;299(5881):371–374.
7. Tsukihara T, Aoyama H, Yamashita E, et al. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. Science 1996;272(5265):1136–1144.
8. Ostermeier C, Harrenga A, Ermler U, Michel H. Structure at 2.7 Å resolution of the *Paracoccus denitrificans* two-subunit cytochrome c oxidase complexed with an antibody FV fragment. Proc. Natl. Acad. Sci. USA 1997;94(20):10547–10553.
9. Zhang Z, Huang L, Shulmeister VM, et al. Electron transfer by domain movement in cytochrome bc_1. Nature 1998;392:677.
10. Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 1996;259(3):393–421.
11. Fyfe PK, Cogdell RJ. Purple bacterial antenna complexes. Curr. Opin. Struct. Biol. 1996;6(4):467–472.
12. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. Science 1997;276(5309): 131–133.
13. Deisenhofer J, Epp O, Sinning I, Michel H. Crystallographic refinement at 2.3 Å resolution and refined model of the photosynthetic reaction centre from *Rhodopseudomonas viridis*. J. Mol. Biol. 1995;246(3):429–457.
14. Ermler U, Fritzsch G, Buchanan SK, Michel H. Structure of the photosynthetic reaction centre from *Rhodobacter sphaeroides* at 2.65 Å resolution: cofactors and protein-cofactor interactions. Structure 1994;2(10):925–936.
15. Doyle DA, Cabral JM, Pfuetzner RA, et al. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. Science 1998;280:69–77.
16. Murzin AG. Structural classification of proteins: new superfamilies. Curr. Opin. Struct. Biol. 1996;6(3):386–394.
17. Breiter DR, Kanost MR, Benning MM, et al. Molecular structure of an apolipoprotein determined at 2.5 Å resolution. Biochemistry 1991;30(3):603–608.
18. Nolde DE, Sobol AG, Pluzhnikov KA, Grishin EV, Arseniev AS. Three-dimensional structure of ectatomin from *Ectatomma tuberculatum* ant venom. J. Biomol. NMR 1995;5(1):1–13.
19. Lawson DM, Artymiuk PJ, Yewdall SJ, et al. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. Nature 1991;349(6309):541–544.
20. Eberle W, Pastore A, Sander C, Rosch P. The structure of ColE1 rop in solution. J. Biomol. NMR 1991;1(1):71–82.
21. Wilson C, Wardell MR, Weisgraber KH, Mahley RW, Agard DA. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science 1991;252(5014):1817–1822.
22. Lederer F, Glatigny A, Bethge PH, Bellamy HD, Matthew FS. Improvement of the 2.5 Å resolution model of cytochrome b562 by redetermining the primary structure and using molecular graphics. J. Mol. Biol. 1981;148(4):427–448.
23. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22(12):2577–2637.
24. Whitley P, Grahn E, Kutay U, Rapoport TA, von Heijne G. A 12-residue-long polyleucine tail is sufficient to anchor synaptobrevin to the endoplasmic reticulum membrane. J. Biol. Chem. 1996;271(13):7583–7586.
25. 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.
26. Gerstein M, Tsai J, Levitt M. The volume of atoms on the protein surface—calculated from simulation, using voronoi polyhedra. J. Mol. Biol. 1995;249(5):955–966.
27. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 1991;24:946–950.
28. Arkin IT, Brunger AT. Statistical analysis of predicted transmembrane α-helices. Biochim. Biophys. Acta In press, 1998.
29. 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.
30. Popot JL, Gerchman SE, Engelman DM. Refolding of bacteriorhodopsin in lipid bilayers. A thermodynamically controlled two-stage process. J. Mol. Biol. 1987;198(4):655–676.
31. Popot JL, Engelman DM. Membrane protein folding and oligomerization: the two-stage model. Biochemistry 1990;29(17):4031–4037.
32. Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM. Sequence specificity in the dimerization of transmembrane alpha-helices. Biochemistry 1992;31(51):12719–12725.
33. Lemmon MA, Treutlein HR, Adams PD, Brunger AT, Engelman DM. A dimerization motif for transmembrane alpha-helices. Nat. Struct. Biol. 1994;1(3):157–163.
34. Adams PD, Arkin IT, Engelman DM, Brunger AT. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. Nat. Struct. Biol. 1995;2(2):154–162.
35. Arkin IT, Adams PD, MacKenzie KR, Lemmon MA, Brünger AT, Engelman DM. Structural organization of the pentameric transmembrane α-helices of phospholamban, a cardiac ion channel. EMBO J. 1994;13(20):4757–4764.
36. MacKenzie KR, Engelman DM. Structure-based prediction of the stability of the transmembrane helix-helix interactions: the sequence dependence of glycophorin A dimerization. Proc. Natl. Acad. Sci. USA Page in press, 1998.
37. Shen ML, Safinya CR, Liang KS, Ruppert AF, Rothschild KJ. Stabilisation of the membrane-protein bacteriorhodopsin to 140°C in 2-dimensional films. Nature 1993;366(6450):48–50.