PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Why are polar residues within the membrane core evolutionary conserved?

Kristoffer Illergård, Anni Kauko, and Arne Elofsson*

Department of Biochemistry and Biophysics, Center for Biomembrane Research, Stockholm bioinformatics Center, Stockholm University, SE-106 91 Stockholm, Sweden

ABSTRACT

Here, we present a study of polar residues within the membrane core of alpha-helical membrane proteins. As expected, polar residues are less frequent in the membrane than expected. Further, most of these residues are buried within the interior of the protein and are only rarely exposed to lipids. However, the polar groups often border internal water filled cavities, even if the rest of the sidechain is buried. A survey of their functional roles in known structures showed that the polar residues are often directly involved in binding of small compounds, especially in channels and transporters, but other functions including proton transfer, catalysis, and selectivity have also been attributed to these proteins. Among the polar residues histidines often interact with prosthetic groups in photosynthetic- and oxidoreductase-related proteins, whereas prolines often are required for conformational changes of the proteins. Indeed, the polar residues in the membrane core are more conserved than other residues in the core, as well as more conserved than polar residues outside the membrane. The reason is twofold; they are often (i) buried in the interior of the protein and (ii) directly involved in the function of the proteins. Finally, a method to identify which polar residues are present within the membrane core directly from protein sequences was developed. Applying the method to the set of all human membrane proteins the prediction indicates that polar residues were most frequent among active transporter proteins and GPCRs, whereas infrequent in families with few transmembrane regions, such as non-GPCR receptors.

Proteins 2011; 79:79–91.

Key words: membrane proteins; polar residues; conservation; accessibility; functional residues.

INTRODUCTION

α-helical transmembrane (TM) proteins constitute about 25% of the proteome in all organisms[1] and they perform many crucial functions, such as signaling, transport across the membrane, energy production, and catalysis. α-helical TM-proteins have been traditionally thought to consist of long hydrophobic α-helices that span the membrane connected through loops. However, recent studies have shown that the structural repertoire of membrane proteins is more complex than was believed only a few years ago.[2] Structural irregularities include kinked TM-helices,[3] interfacial helices,[4] reentrant regions,[5] and coils in the membrane.[6] These irregular structures are often directly associated with the functionality of the membrane proteins. Reentrant regions and membrane coils are common in transporters and channels,[5] while interface helices are common in TM-proteins involved in photosynthesis and respiration.[1]

Water-soluble proteins have a primarily hydrophobic interior and hydrophilic exterior, as it is energetically unfavorable to expose hydrophobic group to the polar aqueous solvent.[7,8] In contrast, the solvent accessible surface of α-helical TM proteins face three distinct environments; a hydrophobic lipid environment inside the membrane, a hydrophilic water environment outside the membrane, and an interface region between these. The energetics is, therefore, more complicated for membrane proteins compared to water-soluble proteins. Early studies of the bacteriorhodopsin structure suggested that membrane proteins would be "inside-out," that is that they consist of a hydrophilic interior and a hydrophobic exterior.[9] However, later studies have shown that the "inside-out" rule is not generally applicable to all α-helical TM proteins.[10] In addition, the existence of the irregular structural features is not consistence with a simple two-state folding process of membrane proteins.

Polar groups that are in contact with the hydrophobic tails of lipids contribute negatively to the overall stability.[11] This is why they are

Additional Supporting Information may be found in online version of the article.

Abbreviations: Charged residues (D, E, K and R); GPCR, G-protein coupled receptor; membrane core, the central region of a transmembrane protein located within 10 Å from the membrane center; RSA, relative accessible surface area; Strongly polar residues (D, E, K, R, H, N, P and Q).

Grant sponsors: Swedish Research Councils, the Foundation for strategic research (SSF) and the EU 6'the Framework program through the EM-BRACE project; Grant number: LSHG-CT-2004-512092; Grant sponsor: Academy of Finland.
Kristoffer Illergard and Anni Kauko contributed equally to this work.
*Correspondence to: Arne Elofsson, Department of Biochemistry and Biophysics, Center for Biomembrane Research, Stockholm bioinformatics Center, Stockholm University, SE-10691 Stockholm, Sweden. E-mail: arne@bioinfo.se.
Received 7 April 2010; Revised 23 July 2010; Accepted 13 August 2010
Published online 23 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22859

rare within the membrane core.[12] In general, polar groups that are removed from the contact with water are stabilized by interactions with other polar groups.[13] Therefore, within the membrane core, polar groups do only rarely interact directly with the hydrophobic lipid tails. Instead, they interact with other polar groups in the protein, or line internal water filled cavities. Further, polar groups at the end of the polar sidechains can snorkel out toward the phospholipid head-group.[4] Polar residues are also unfavourable for insertion of individual helices into the membrane by the translocon machinery.[14,15]

The use of multiple sequence alignments to identify conserved residues has provided many important clues about the function of inividual proteins. Buried residues are in general more conserved than exposed residues, and it was recently observed that one reason why membrane proteins appear to be more conserved than soluble proteins is due to the fact that they bury a larger fraction of their surface area.[16] However, increased conservation is also an important indicator of functional importance as previously observed for coil residues within the membrane core.[6]

In this study, we have tried to shed some light on the roles of we have characterized strongly polar residues (D, E, K, R, H, N, P, and Q) within the membrane core. Initially, the incidence and evolutionary conservation of these residues in different environments was analyzed to establish that they are rare and evolutionary conserved. To gain insight into the increased conservation, we assess their structural and functional roles by characterizing the microenvironments and conduct an investigation of the literature. Finally, a sequence-based method to identify polar residues within the membrane core was developed and applied to all human membrane proteins and the distribution of these residues in different classes of membrane proteins was analyzed.

RESULTS AND DISCUSSION

In this study, strongly polar residues located within the membrane core in α-helical TM proteins were analyzed from a structural, functional, and evolutionary perspective. The aim was to gain insight into their role within the membrane core and to investigate the generality of earlier observations from studies of individual proteins. Here, we have analyzed these residues in a large high quality dataset of membrane proteins of known structures. Finally, we use the collected information to predict polar residues in all human membrane proteins and to speculate about their structural and functional roles.

Strongly polar residues in the membrane core are rare, buried and conserved

A comparison of the amino acid frequency at different distances from the membrane center in the dataset confirmed that polar residues are less frequent within the core of the membrane bilayer than outside the membrane, Figure 1(A). It was found that the strongly polar residues (D, E, K, R, H, N, P, and Q) constitute only 9 % of the residues within the membrane bilayer compared to 41 % outside the membrane. Further, the observed difference in preference for amino acids between inside and outside the membrane is strongly correlated with the previously defined biological hydrophobicity scale.[15] This scale describes the experimentally obtained free energy cost of translocon recognition from the ΔG-predictor[15] (R^2 = 0.88) Figure 1(B). Although the correlation is high, the biological scale suggests a higher free energy cost for insertion of all residues than the scale based on the residue frequencies. This is in agreement with the fact that it is necessary to set the ΔG cutoff higher than expected (zero) when using the biological hydrophobicity scale to predict regions that after folding will become TM-regions.[17]

The next feature to investigate was the substitution rate for all sites. The analysis shows that polar residues are clearly more conserved within the membrane than outside, Figure 1(C). These residues also show the largest variation in substitution rate across the membrane. It has been noted before that polar residues are more conserved within the membrane.[18] In addition, we found that the difference in substitution rates between the membrane and non membrane regions of the amino acids is correlated with the biological hydrophobicity scale (R^2 = 0.70), Figure 1(D). One notable outlier to the observed relationship between hydrophobicity and conservation is cysteine, which is more conserved than expected outside the membrane, Figure 1(D). The behavior of cysteine is likely due to that many cysteines form conserved disulphide bonds outside the membrane, whereas no disulphide bonds are present in the membrane core in any of the membrane proteins studied here.

The third feature to study was accessibility. It was found that the average accessibility of polar residues varies strongly with the position relative to the membrane, Figure 1(E). Outside the membrane polar residues are more likely to be exposed, whereas inside the membrane they are more likely to be buried, Figure 1(E).

In general, conservation is an indication of structural or functional importance. Although these features are intimately related, we try to identify primarily functionally important residues by detecting residues that are more conserved than other residues within similar microenvironments. The most prominent structural factor affecting substitution rate is the exposure level (accessibility), so that residues buried in the protein interior are more conserved than solvent exposed residues.[16] Therefore, it is possible that the increased conservation of polar residues in the membrane core is related to the preference for them to be buried. However, by comparing sites of similar exposure level, we found that strongly polar residues are actually more conserved than other resi-

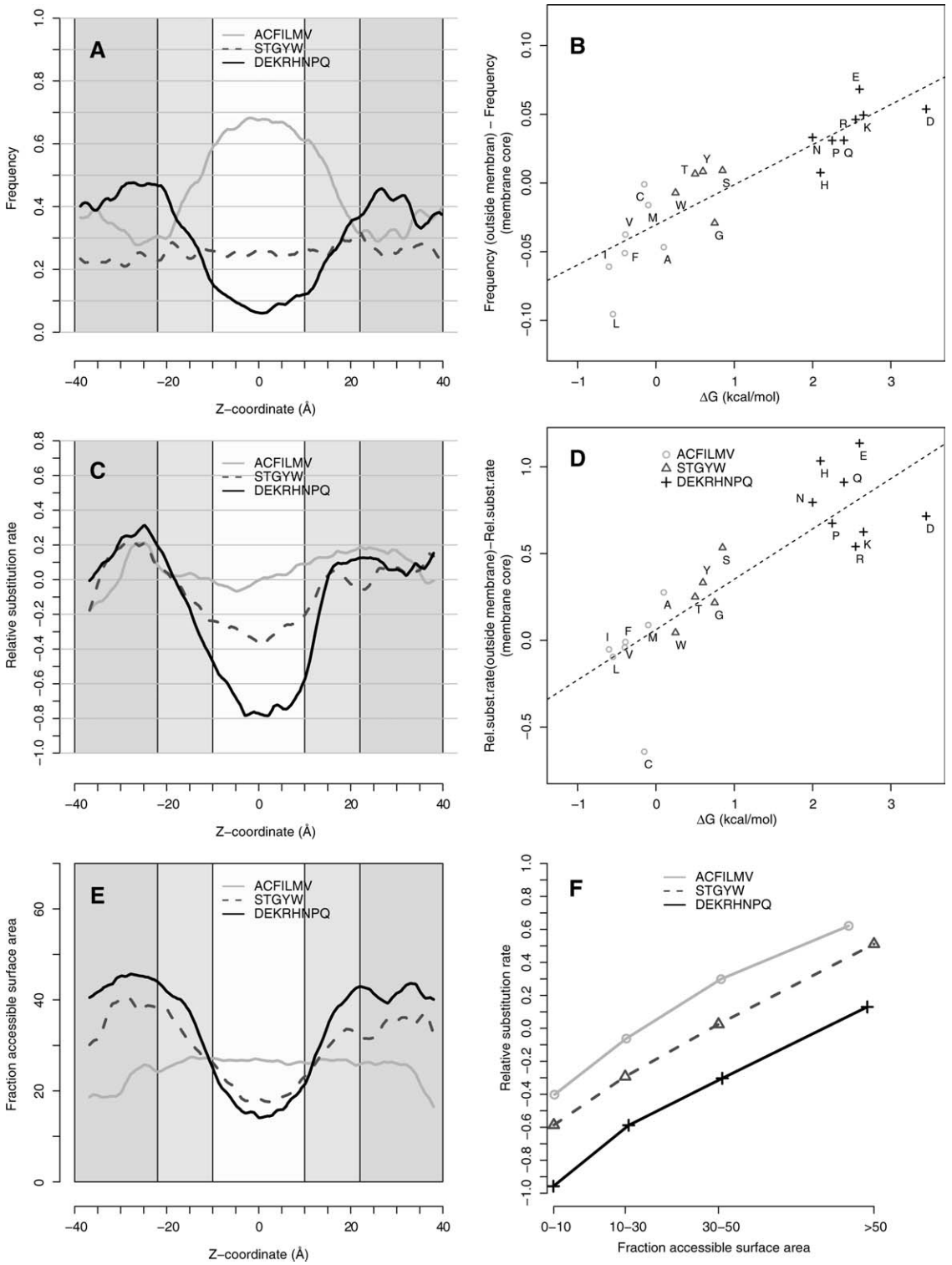

Figure 1

Properties of different residues in membrane proteins. **A**, **C**, and **E**: These show the distribution of amino acids, substitution rates, and relative surface accessibility across the membrane, respectively. **B**: The difference in frequency between the outside and inside of the membrane for different amino acids and **D**: the difference in substitution rates at sites positioned outside and inside of the membranes plotted against biological hydrophobicity scale. **F**: The substitution rate at comparable accessibility levels for different amino acid groups within the central membrane region are shown. In all figures, the coloring differentiates between polar, intermediate, and hydrophobic residues.

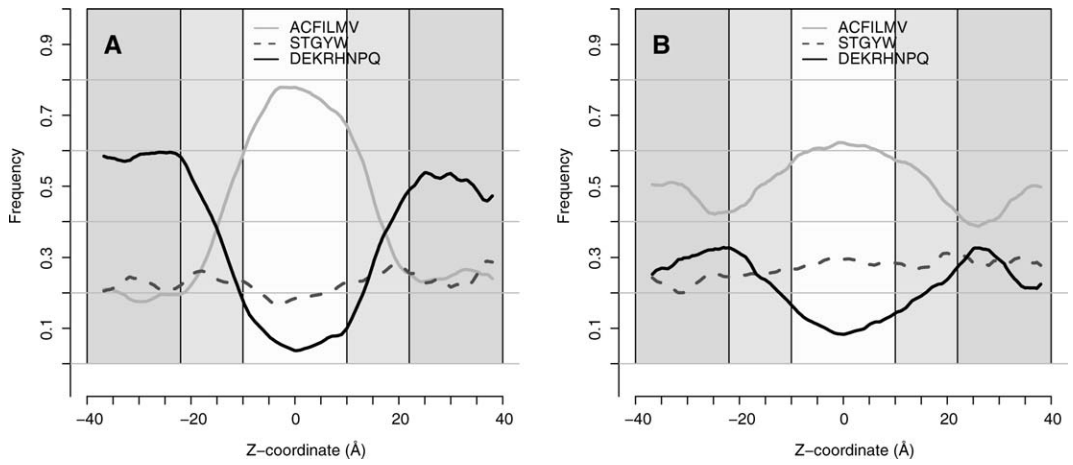

Figure 2

The distribution of amino acids across the membrane. **A**: The subset of residues that have accessibility values larger than 25% (solvent accessible), whereas in **B**: the subset of residues that have accessibility value lower or equal to 25% (solvent inaccessible) can be seen. The sum of the values in the two figures would result in Figure 1A. In all figures, the coloring differentiates between polar, intermediate, and hydrophobic residues.

dues in the membrane, even when taking the accessibility into account, Figure 1(F). This indicates that other factors, such as direct functional importance, also contribute to the strong conservation of polar residues. The importance suggested from the conservation is also supported by an over-representation of these residues in disease causing mutations.[19]

Strongly polar residues within the membrane core often interacts with water

The findings above suggest that the incidence of residues at different distances from the membrane core strongly depend on their polarity, Figure 1(A, B). However, buried residues in the protein interior show much less dependency on the distance from the membrane center, than residues that are solvent accessible, Figure 2(A, B). The frequency of strongly polar residues among the buried residues only decrease from 26% outside the membrane to 11% inside the membrane, whereas the corresponding fraction among exposed residues decrease from 59% to 6%. The tendency of small differences in amino acid distributions for the protein interior has been noted before.[20]

Further, the interaction patterns of strongly polar sidechain groups were characterized from the structures, Figure 3. The fraction of polar atoms in contact with water molecules or putative water molecules in internal cavities was only slightly lower inside than outside of the membrane (34% and 52%, respectively, P = 0.16, by Fischer's exact test). Accordingly, the fraction of polar atoms interacting with other polar groups from the protein is quite similar inside and outside the membranes (54% and 47%, respectively, P = 0.12, by Fischer's exact test), Figure 3. Thus, even though the strongly polar sidechains are mostly buried within the membrane core, the polar atoms are often located near water containing cavities. The presence of water around polar groups is in line with what has been observed in simulations of amino-acid analogs within the hydrophobic membrane core.[21]

In addition to the strongly polar residues discussed in the rest of the article, serine and threonine also contain polar groups. In contrast to the strongly polar residues, they preferably interact with the backbone (86 % and 53 %, respectively, $P < 10^{-9}$ by Fischer's exact test), as previously observed in soluble proteins.[22] To create shared hydrogen bonds within the α-helix, these sidechains often point toward the N-termini.[23] Shared hydrogen bonds probably explain, why Ser and Thr belong to the category "weakly polar" when defined by insertion efficiency, though the hydroxyl group itself is strongly polar.

A method to identify strongly polar residues in the membrane core

A limitation of studying only membrane proteins of known structures is that many functional classes of TM-proteins, like receptors, are only sparsely represented in such datasets. To overcome this, we examined the possibility of developing a method to identify polar residues within the membrane core directly from the protein sequence.

Initially, it was found that by predicting the distance from the membrane center by Zpred[24,25] and setting the cutoff to 10 Å, it is possible to accurately identify a large fraction of all polar residues in the membrane core. Among structurally determined proteins, Zpred predicts that the core contains 7.0% polar residues, compared to

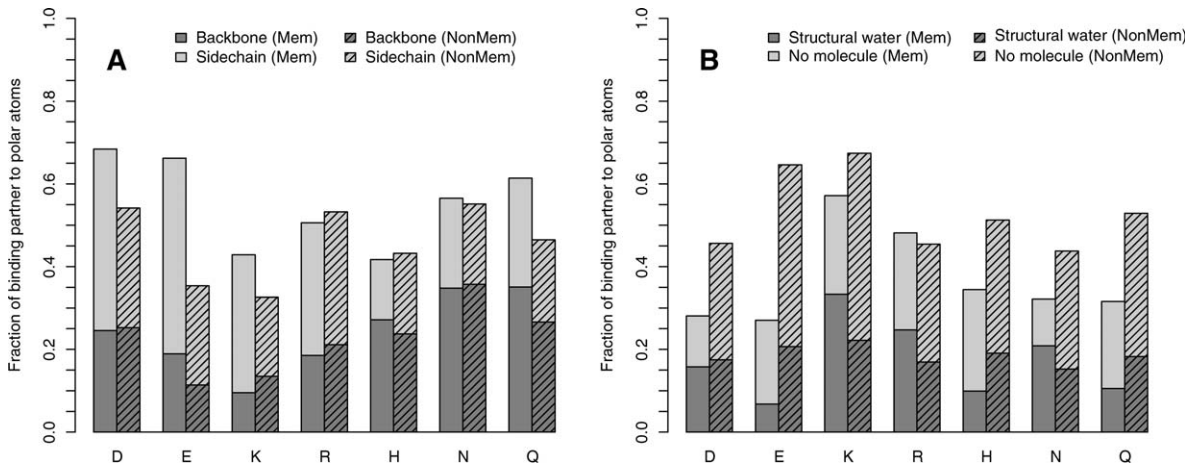

Figure 3

The distribution of the environment around the polar groups inside (Mem) and outside (NonMem) the membrane. The environment around a polar group is divided into five groups; backbone, sidechain, prosthetic group, structural water, or "no molecule". **A**: This shows the fraction of groups involved in hydrogen binding to protein backbone or protein side-chain. **B**: This shows the fraction of groups interacting with structural water or not having any polar interaction partner ("No molecule").

the observed 9.5% and, additionally, 58% of the polar residues in the core were correctly identified and the corresponding accuracy was 95% (Matthews Correlation Coefficient = 0.64), see Table I.

Thereafter, the method for predicting the polar residues in the membrane core was applied to all human membrane proteins. The estimated overall frequency was found to be 7.6%. It should be remembered that quite a few strongly polar residues in the membrane core are probably missed by the predictions. However, it is encouraging that the identified residues are generally predicted to have similar properties to the ones in the dataset of known structures. For instance, the polar residues that are predicted to be in the core or outside the membrane, respectively, have average frequency, solvent accessibility, and substitution rate that correspond quite well to the ones in the structural dataset, Table I. Thus, the predictions also point to the fact that polar residues are less frequent in the core compared to the outside of the membrane, and that they are evolutionary conserved as well as buried in the protein interior.

After establishing that the method can identify polar residues that have similar properties to the ones in the dataset of known structures, we wanted to investigate the structural and functional properties of these residues. First, we wanted to investigate whether the polar residues in the core regions are evenly distributed in the sequence. This seems not to be the case. The first TM-helix are predicted to contain on average fewer polar residues in the core than later TM-helices (5% vs. 7.6%), Figure 4(A). This finding is in agreement with the observation that the N-terminal TM-region is required to be more hydrophobic than subsequent regions to be recognized by the translocon machinery.[17] The predicted set of polar residues in the membrane indicates that proteins with few transmembrane regions contain a lower number of polar residues, see Figure 4(B). This is in contrast to what has been shown in spherical water-soluble proteins,[27] but could be expected as the volume to surface area decreases with size and polar residues are rare at the surface. However, a larger dataset of membrane proteins of known structures is needed to finally verify these observations.

Fraction of strongly polar residues in different proteins

To gain further insight in the functional role of strongly polar residues in the membrane core, we present here a first large-scale comprehensive study of their frequency in different protein families, see Tables II and III.

Table I

Estimated Frequency, Accessibility, and Normalized Substitution Rate of Polar Residues Position Inside and Outside the Membrane in Membrane Proteins

Dataset		Frequency		Accessibility		Subst.rate	
Dataset	Estimation	Inside	Outside	Inside	Outside	Inside	Outside
Human	pred	7.6 %	46.0 %	20.4 %	44.0 %	−0.59	0.12
OPM	pred	7.0 %	45.4 %	17.8 %	41.6 %	−0.66	0.20
OPM	struct	9.5 %	41.2 %	20.5 %	41.0 %	−0.67	0.18

The residues found <=10 Å from the membrane center are regarded as "inside" the membrane, and residues >22 Åfrom the the membrane center as "outside" the membrane. The distances are predicted by Zpred, top two lines, or determined from the *Z*-coordinate from the structures oriented as in OPM, third line. The accessibility of each protein is determined with Naccess or predicted using MPRAP.[26] The relative substitution rate is measured by rate4site. Two datasets are used, either all human membrane proteins (Human) or the homology reduced dataset of membrane proteins of known structures (OPM).

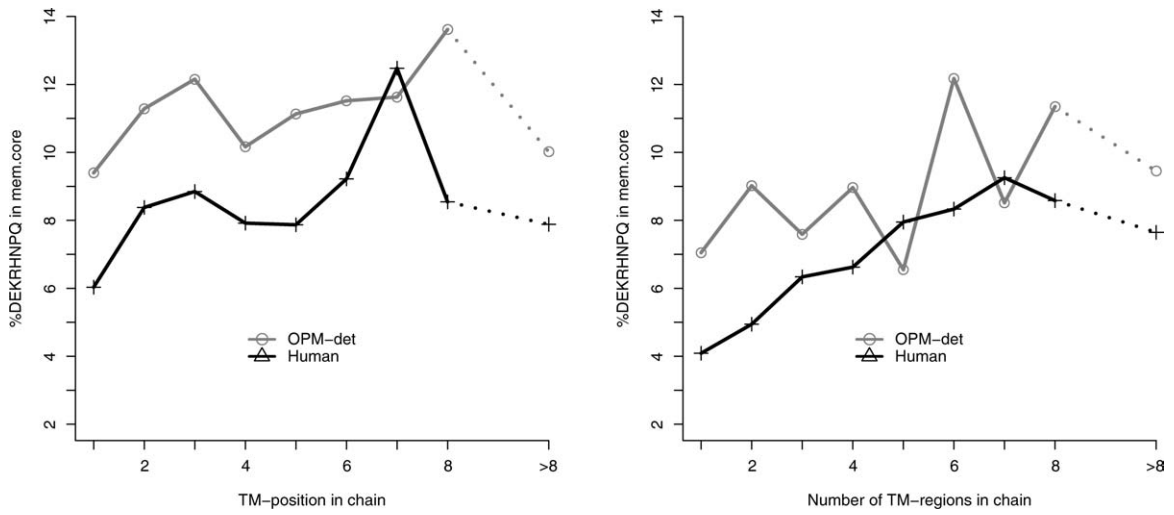

Figure 4

The fraction of polar residues (D, E, K, R, H, N, P, and Q) within the membrane core region. **A**: Thus shows the distribution as a function of transmembrane segment position and **B**: as a function of number of transmembrane regions in the protein chain. The two datasets are all human membrane proteins (Human) and the homology reduced dataset with membrane proteins of known structures (OPM).

Among the proteins of known structure, the highest fractions of strongly polar residues were found in transporters (10.1%) and receptors (9.5 %), Table II. From the analysis of the larger dataset of all human membrane proteins, the high frequency in these groups could be traced to G-protein Coupled Receptors (GPCRs) and active transporters, Table III.

Our analysis from the predicted set of polar residues in the human membrane proteins shows that the significant enrichment of polar residues in GPCRs compared to all others is largely constituted by histidines, asparagines, and prolines, Table III. The function of GPCRs is to recognize molecules outside the cell and activate signal transduction pathways and cellular responses inside the cell.[28,29] The GPCRs of known structures have two clusters of polar residues in the membrane core, which are related to these functions. The first is the binding site for the ligand located in a deep crevice. The second cluster is likely involved in the conformational changes that occur upon activation and deactivation. Additionally, several prolines probably also play an important role for the flexibility.[3]

It was also found that active transporters in the human dataset are enriched in polar residues, especially in prolines and glutamines, Table III. This is in agreement with our observations from the structural dataset. Active transporters are in humans dominated by two protein families, the ABC transporters[30] and P-type ATPases.[31]

Table II

Incidence of Polar Residues (D, E, K, R, H, N, P, and Q) within the Membrane Core of Proteins of Known Structure

Main group	Sub group	#Chains	#TMs	%D	%E	%K	%R	%H	%N	%P	%Q	%All
All		85	4.4	0.8	0.7	0.6	0.9	1.6	1.5	2.5	0.8	9.4
Enzymes	All	29	3.0	0.8	0.6	0.5	0.7	2.3	1.2	2.0	0.6	8.8
	Others	6	4.2	0.7	0.7	0.5	1.2	1.9	1.9	1.7	0.5	9.0
	Oxidoreductase(Respiration)	23	2.7	0.8	0.6	0.5	0.5	2.5	0.9	2.2	0.6	8.7
Photosynthesis	All	18	3.3	0.0	0.2	0.1	0.5	**4.7**	0.8	1.6	0.3	8.3
Transporters	All	32	6.7	1.0	0.9	0.8	1.0	**0.5**	1.7	3.0	1.0	10.1
	Active transporters	11	7.0	0.8	1.0	0.6	0.8	0.9	1.5	3.4	0.8	10.0
	Channels	11	3.1	0.7	0.9	0.3	1.0	**0.3**	1.7	2.1	1.2	8.2
	Potential driven	9	10.9	1.3	0.9	1.1	1.2	0.3	1.8	3.1	1.1	10.9
	Unclassified	1	5.0	1.2	0.0	1.2	1.2	0.0	3.8	2.5	0.0	10.0
Receptor	All	2	4.0	2.6	0.9	0.0	0.0	0.0	3.4	2.6	0.0	9.5
	GPCR	1	7.0	1.9	1.0	0.0	0.0	0.0	3.9	2.9	0.0	9.7
	Others	1	1.0	7.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7
Miscellaneous	All	4	1.2	0.0	0.0	0.0	0.0	0.0	0.0	1.7	1.7	3.3

Number of polypeptide chains and transmembrane regions (TMs), as well as the percentage of polar residues of different types is shown for proteins classified into different functional groups. All values that are significantly ($P < 10^{-5}$) diverging from the average are marked in bold.

Table III
Predicted Incidence of Polar Residues (D, E, K, R, H, N, P and Q) in Membrane Core of Human Membrane Proteins

Main group	Sub group	#Chains	#TMs	%D	%E	%K	%R	%H	%N	%P	%Q	%All
All		5737	3.8	0.7	0.7	0.4	0.6	0.8	1.5	2.3	0.9	7.7
Enzymes	All	532	2.7	0.4	0.6	0.5	0.7	0.9	0.9	2.1	0.8	7.1
	Hydrolases	181	2.1	0.5	0.3	0.3	0.9	0.5	0.5	2.4	0.7	6.1
	Isomerases	8	2.6	0.7	0.7	0.4	0.7	1.5	1.1	1.8	0.0	6.9
	Ligases	7	1.3	0.6	1.3	0.6	0.6	1.3	1.3	1.9	1.3	9.0
	Lyases	17	6.9	0.3	0.9	0.6	0.3	1.0	1.3	1.0	1.1	6.5
	Oxidoreductases	126	2.7	0.4	0.9	0.6	0.8	1.1	1.3	2.6	0.6	8.2
	Transferases	193	2.8	0.3	0.5	0.5	0.7	1.1	0.9	1.9	1.1	7.1
Transporters	All	817	7.7	0.7	0.8	**0.4**	0.7	**0.5**	1.5	2.2	**1.0**	7.7
	Active transporters	81	9.8	0.7	0.8	0.5	0.6	0.6	1.4	**2.9**	**1.6**	**9.0**
	Auxiliary transport unit	42	3.6	0.1	0.7	0.4	0.5	0.5	0.9	1.4	1.0	5.5
	Channels	246	6.4	**1.1**	**1.1**	0.5	0.7	**0.4**	**1.9**	**1.5**	0.7	7.9
	Potential driven (SLC)	393	9.1	**0.5**	0.6	0.4	0.7	**0.4**	1.3	2.4	1.0	7.3
	Others	55	3.9	0.3	0.4	0.7	1.0	1.0	0.7	1.7	1.3	7.0
Receptors	All	1350	5.0	**0.9**	0.7	**0.3**	**0.5**	**1.1**	**2.1**	**2.6**	0.8	**8.9**
	GPCR	899	6.7	**1.0**	0.7	**0.3**	**0.5**	**1.2**	**2.3**	**2.6**	0.8	**9.4**
	IG	149	1.2	**0.1**	0.3	0.3	0.4	0.2	**0.2**	2.5	0.2	**4.3**
	Kinase	72	1.7	1.4	0.9	0.1	0.5	0.2	**0.1**	2.1	0.5	5.8
	SCAR	63	1.3	0.0	0.2	0.3	0.5	0.1	0.3	1.0	0.2	**2.7**
	Others	167	2.1	0.4	0.3	0.4	0.7	0.7	**0.5**	2.2	0.6	**5.7**
Miscellaneous	All	684	2.5	**0.1**	0.7	0.3	0.6	0.7	**1.0**	**1.6**	0.8	**5.8**
	Ligand	57	1.6	0.0	0.1	0.1	0.7	0.1	0.9	1.6	0.3	**3.8**
	Other	271	2.5	**0.1**	0.8	0.3	0.6	0.6	1.0	**1.6**	1.0	**6.0**
	Structural And Adhesion	187	2.1	**0.1**	0.7	0.2	**1.2**	**0.2**	**0.6**	**1.3**	0.9	**5.3**
	Unknown	169	3.2	**0.2**	0.7	0.4	0.3	1.2	1.1	1.7	0.8	**6.4**
Unclassified	All	2354	2.3	**0.5**	0.6	0.4	0.7	0.8	**0.9**	2.2	0.8	**6.7**

Average number of polypeptide chains and transmembrane regions (TMs), as well as the fraction of polar residues is shown for different functional groups. All values that are significantly ($P < 10^{-5}$) diverging from the average are marked in bold.

Both groups convert energy by ATP hydrolysis and undergo large conformational changes during their reaction cycle. The structurally variable ABC transporters transport a wide variety of substances, whereas P-type ATPases transport ions. In both cases, prolines might be important for conformational changes in the transportation cycle, and polar residues are common at binding sites or internal cavities.

In summary, the high fraction of polar residues in some groups of proteins can partly be related to functional need of the proteins in those groups. Another factor related to the incidence of polar residues in the membrane core is the number of TM regions (see above). For example, the non-GPCR receptor subgroups contain significantly less polar residues than GPCRs. Often these proteins only contain a single TM region.

Analysis of functional residues

To gain a more detailed insight into the functional roles of individual strongly polar residues within the membrane core, we classified the environment and the functional role of all these residues based on literature and structures. Although this survey is certainly not complete, nor unbiased, it might nonetheless provide some ideas about the functional roles of strongly polar residues. It could also give an indication of the generality of earlier observations.

In transporter proteins, most of the functionally annotated residues are involved in ligand binding,[32–34] but various other annotations were also found, including gating,[35] flexibility,[36] sensors and proton transfer,[37] see Table IV. The role of Proline is often described to provide flexibility for conformational changes during the transport cycle of polar substances through the membrane. Most functionally annotated polar residues are buried, as in the voltage sensor domain of the potassium channel, Figure 5(A). However, some were partly exposed to large internal channels, Figure 5(B) and in a majority of these residues the actual polar group was exposed to small cavities, see Figure 5(C).

A relatively large fraction of proteins in the structural dataset are involved in photosynthesis or a part of the respiratory electron transport chain. These proteins have distinct structural and functional features compared to other membrane proteins. To some extent, these structures can be described as cages whose main function is to orient the prosthetic groups correctly, and in many cases, to power a proton pump using the energy from oxidoreduction or photosynthesis. The total number of histidines in membrane core is significantly higher in photosynthetic proteins than in other membrane proteins (4.7% vs. 1.6%, Fisher's exact test $P < 4 * 10^{-11}$) and histidines are overrepresented also in oxidoreductases from respiration, see Table II. Accordingly, a large number of functionally important histidines that bind heme or chlorophyll exist in these proteins,[43,44,53] Figure 5(D).

Table IV
Functionally Important (according to literature annotations) Polar Residues (D,E,K,R,H,N,Q,P) within the Membrane Core and Depicted by Authors to be Directly Involved in the Function

Group	Protein name	PDB id	Residue function	Residues
Oxidoreductases from respiration	Formate dehydrogenase[38]	1kqfC	Heme binding	H57, H155
	Cytochrome bc1[39]	1pp9P	Heme binding	**H97**, H196
			Substrate binding	**D228**
	Respiratory Nitrate Reductase[40]	1q16C	Heme binding	H56, **H66**, H187
	Mitoch. Cytochrome c oxidase[41,42]	1v55A	Heme binding	H61, H376, H378
			Cu_B binding/Proton pathway	H240
			Proton pathway	**E242**, **K319**, **H413**
	Succinate dehydrogenase[43]	1yq3C	Heme binding	H98
		1yq3D	Heme binding	**H46**
	Fumarate reductase[44,45]	2bs2C	Heme binding	H44, **H143**, H182
			Coupling	**E180**
Other enzymes	Glutathione S-transferase 1[46]	2h8aA	Catalytic	**E80**
	Protease glpG[47,48]	2ic8A	Catalytic	**H150**, **H254**
			Gating	**H145**, **N154**
	Metallo protease S2P[49]	3b4rB	Catalytic	**H54**, **E55**, **N140**, **D148**
Photosynthesis	Photosystem I[50]	1jb0A	Chlorophyll binding	**N604**, H680
			Chlorophyll binding	H76, H79, H199, H200, H313, H396 **H397**, H443, H454, H540, H547, H734
		1jb0L	Chlorophyll binding	E49, H54
	Light harvesting complex LH2[51]	1nkzA	Chlorophyll binding	H31
		1nkzB	Chlorophyll binding	H30
	Photosynthetic reaction center[52]	1rzhL	Chlorophyll binding	**H173**
		1rzhM	Chlorophyll binding	**H202**
	Photosystem II[53]	2axtA	Chlorophyll binding	H118, H198
		2axtB	Chlorophyll binding	H23, **H26**, H100, H142, H201,H202, H455, H466
		2axtC	Chlorophyll binding	H53, H56, H118, H164, **H251**, H430, H441
		2axtD	Chlorophyll binding	H117, H197
		2axtE	Heme binding	H23
	Light harvesting complex II[54]	2bhwA	Chlorophyll binding	H68, **E139**, N183
Active transport	Bacteriorhodopsin[55]	1m0lA	Proton pathway	**R82**, **D85**, **D212**
			Retinal binding	**K216**
	Preprotein translocase SecY[56]	1rh5A	Gating	E122, **N268**
	P-type Ca^{2+} ATPase[34,57]	1wpgA	Ion binding	**N768**, E771, **N796**, **D800**, **D908**
			Gating/Ion binding	**E309**
			Flexibility	P308
	V-type Na^+ ATPase[58]	2bl2A	Ion binding	Q65, *Q110, E139*
	Molybdate transp. ModC (ABC)[59]	2onkC	Gating	P94, P98, **R195**, **E199**
	Maltose transporter MalF (ABC)[60]	2r6gF	Substrate binding	**N376**, **N437**
Potential driven transport	Mitoch. ATP/ADP carrier[32]	1okcA	Substrate binding	K22, R79, R187, R279
			Gating	**R137**
	Glycerol-3-P transporter GlpT[61–63]	1pw4A	Substrate binding	**R45**, K80, **H165**, **R269**
			Flexibility	P172 P364
	Na^+/H^+ antiporter I NhaA[64]	1zcdA	Ion binding	**D163**, **D164**
			Flexibility	P129, P265
	Leu transporter LeuT[35]	2a65A	Ion binding	**N27**, **N286**
			Gating	**R30**, Q250, E290, D404
	Lactose permease[36]	2cfqA	Proton pathway	R302, H322, **E325**
			Coupling	**E269**
			Flexibility	P28, P123, P327
	Multidrug transporter EmrD[65]	2gfpA	Proton pathway	**E227**
	Multidrug transporter AcrB[37]	2gifB	Proton pathway	**D407**, **D408**, **K940**
	Glu transporter homolog[33]	2nwlA	Substrate binding	**D394**, **R397**, **N401**
			Ion binding	**D405**
Channels	Clc chloride channel[66]	1otsA	Gating	**E148**
	Aquaporin Aqpm[67,68]	2f2bA	Selectivity	**N82**, **N199**,**R202**
	Acid sensing ion channel[69,70]	2qtsB	Ion binding	D433
	Voltage gated K^+ channel[71]	2r9rB	Voltage sensing	**R293**, R296, **R299**, **K302**
	Cor A Mg^{2+} transporter[72]	2bbjA	Flexibility	P303
Unclassified transp.	Mg^{2+} transporter MgtE[73]	2yvxA	Ion binding	**N329**, **N332**, **D432**
Receptors	Beta2-adrenergic receptor (GPCR)[28,74]	2rh1A	Substrate binding	**D113**, N312
			Flexibility/Regulation	**N51**,**D79**,**N322**
			Flexibility	P88, P211, P288
	T cell receptor dimeris. domain[75]	2hacA	Dimerisation	*D6*

Buried residues (accessibility < 25 %) that have polar group next to the internal cavity are shown as bold. Buried residues with polar group next to membrane are shown as italic. Exposed residues (accessibility > 25 %) are underlined. Histidines are highlighted in gray.

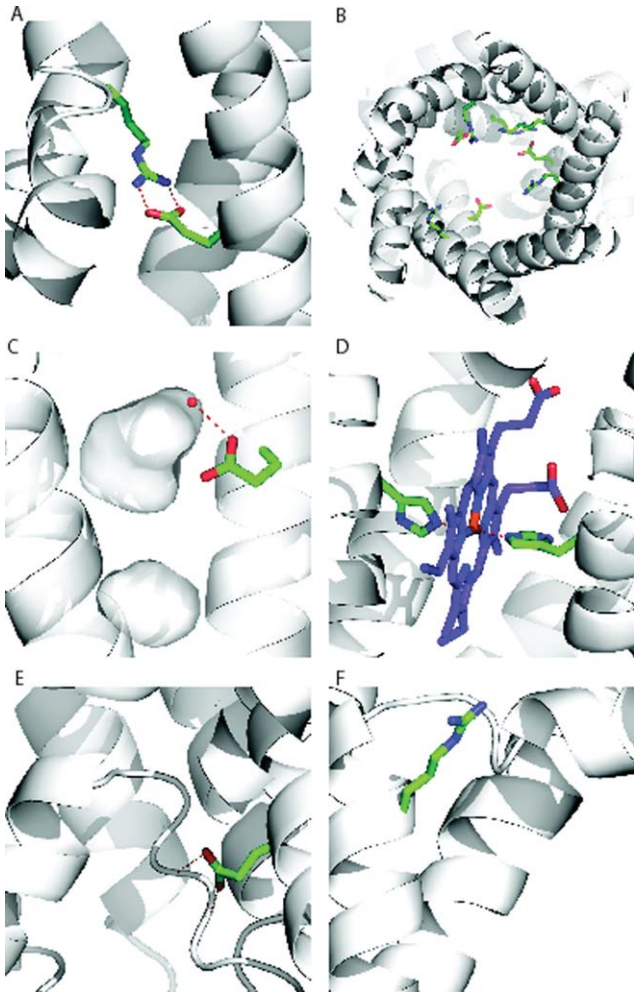

Figure 5

A: A salt bridge in the voltage sensor domain of the potassium channel (2r9rB) between two charged residues (E226, R299).[71] **B**: The large functional channel of the ATP carrier protein (1okcA) with many charged residues (K22, E29, R79, D134, D137, R187, D231, R279) lining the inside of the channel entrance.[32] **C**: A buried glutamate (E908) in calcium ATPase (1wpgA) with carboxylate group exposed to a small cavity.[34]; **D**: Two Histidines (H66, H187) in the respiratory nitrate reductase (1q16C) binding a co-factor.[40] **E**: A carboxylate group in aquaporin (2f2bA) in a glutamate (D11) that serves to stabilize a coil segment that lines the functional pore.[67] **F**: A lipid exposed arginine (R63) in calcium ATPase (1wpgA) with a charged group snorkeling to the water–lipid interface region.[34]

Finally, it should not be forgotten that residues that appear not to be directly involved in the function might have important structural roles. They may, for example, stabilize functional residues, Figure 5(E), form polar interiors of the channels, Figure 5(B), or anchor TM regions to the membrane with long charged sidechains that can snorkel to the water–lipid interface, Figure 5(F). Further, helix-breaking prolines can bring mainchain polar groups into use,[6] as for example in aquaporin.[67,68]

In addition to the literature analysis of functionally annotated polar residues, a semi-manual classification of the microenvironment, interactions, and functional importance was performed for all 545 strongly polar residues in our dataset. The full data are available as Supporting Information and a summary is presented in Table V. All residues were placed in one out of six microenvironments: exposed to cavity; buried, polar group toward cavity; buried; buried, polar group toward membrane; exposed to the membrane; and polar group near the interface. In addition, the functional annotation from the literature analysis and a classification if the polar group had a proper binding partner present was included in the analysis. The roles of Histidines and Prolines is quite different from the other residues, and it was found that they had other environmental preferences than the other residues.

As mentioned above, most of the polar groups are buried in the protein and, in particular, many (60%) of the polar groups in D, E, K, R, Q, N are lining an internal cavity, and ~40% of such residues have been annotated to be directly involved in function. The cavity exposed polar residues mainly create surfaces for water filled cavities, and about one fifth of these have been annotated to be functional. In contrast, many of the histidines are fully buried as they are in contact with prosthetic groups. It is also clear that very high fraction (~70%) of the histidines has been annotated in the literature as functionally important. A large fraction of the kinks introduced by prolines induce membrane exposed main chain carbonyls. Most probably the reason for this is that flexible hinges are needed in all parts of the structure and they cannot all be hidden from the membrane core. Finally, membrane exposed polar groups are rare (<10%) and only seldom annotated to be functionally important (only three examples were found). However, it would be surpising if such destabilising residues exist without a clear functional role, so their roles remain unexplained.

CONCLUSIONS

Here, we have characterized the role of polar residues within α-helical membrane proteins. These residues constitute as much as 9% of all residues in the membrane core of alpha-helical membrane protein structures, and a similar fraction is predicted to exist within all membrane proteins of the human genome. Frequently, the polar groups form hydrogen bonds with water or border internal cavities. In general, it can be said that the environment that surrounds strongly polar residues in the membrane core resembles the environment that surround strongly polar residues in the interior of soluble proteins. Further, it was found that the N-terminal TM-region contains on average less polar residues than subsequent regions. Also, predictions concluded that G-Protein Coupled Receptors and active transporters contain more polar residues than other membrane proteins. This could partly be explained by their functional importance in these protein groups.

Table V
Microenvironments for Polar Groups within the Membrane Core

Environment	D	E	K	R	N	Q	H	P	All
Number of residues in each environment									
All	46	42	35	49	86	47	91	149	545
Exposed to cavity	12	3	8	7	12	5	4	4	55
Buried, polar group toward cavity	23	21	10	21	46	18	25	36	200
Buried, polar group buried	3	10	4	2	8	12	50	31	120
Buried, polar group toward membrane	3	5	0	1	2	1	3	23	38
Exposed to membrane	0	1	2	4	6	3	5	32	53
Near the interface	5	2	11	14	12	6	4	23	77
Percentage functional residues									
All	37%	40%	17%	31%	22%	6%	68%	9%	28%
Exposed to cavity	25%	33%	25%	43%	8%	0%	25%	25%	22%
Buried, polar group toward cavity	57%	52%	40%	48%	37%	0%	60%	17%	38%
Buried, polar group buried	0%	40%	0%	100%	12%	17%	92%	6%	48%
Buried, polar group toward membrane	33%	20%	0%	0%	0%	100%	0%	13%	16%
Exposed to membrane	0%	0%	0%	0%	0%	0%	0%	6%	4%
Near the interface	0%	0%	0%	0%	0%	0%	0%	0%	0%
Percentage residues without proper binding partners									
All	7%	7%	6%	4%	10%	19%	9%	46%	19%
Exposed to cavity	0%	0%	0%	0%	0%	0%	0%	0%	0%
Buried, polar group toward cavity	0%	5%	10%	0%	7%	6%	4%	25%	8%
Buried, polar group buried	33%	0%	0%	0%	25%	25%	2%	55%	20%
Buried, polar group toward membrane	67%	20%	0%	0%	0%	0%	67%	74%	58%
Exposed to membrane	0%	100%	50%	50%	67%	100%	80%	75%	74%
Near the interface	0%	0%	0%	0%	0%	0%	0%	4%	1%

The top part of the table lists the number of residues of each type in each microenvironment. The second section lists the fraction of residues in of a particular type that was found to be functional in this microenvironment. The third section lists the fraction of residues without a proper binding partner present in the structure. To take the focus away from fractions that are based on less than 10 data points these are shown in a smaller font. It should be noted that assignment of functions is dependent from authors of original papers. In particularly, many conserved prolines probably playing role for flexibility were not noted in literature and thus not included here. In contrast, residues binding prosthetic groups (mainly histidines) could be reliably picked from structures and thus they have most complete classification. This biases the statistics for functional importance.

Polar residues are unfavorable in the hydrophobic membrane core.[14] Further, functionally important residues can contribute unfavorably to the stability of the native state[76,77] and are evolutionary conserved. In agreement with these ideas, polar residues in the core are in general more conserved than other residues. The conservation can partly be explained by their tendency to be buried in the protein interior but is also in many cases due to do the direct involvement in the function of the protein. The polar residues perform various functions, including binding, transport, catalysis, gating, and providing flexibility. Heme and chlorophyll are mainly bound by histidines, while flexibility is often provided by prolines. We have previously shown that deep core coils in membrane proteins are irregular structures within the membrane that frequently are coupled with functional importance.[6] Polar residues in the membrane core are examples of structural elements that are important for the function of alpha-helical membrane proteins.

METHODS

Dataset

136 α-helical TM protein structures containing 601 polypeptide chains with TM segments from OPM[78] was downloaded in April 2008. Poly-alanine chains, theoretical models and obsolete entries (as defined by pdb) were excluded. In addition, poor quality structures were removed from this dataset. These were: fragments of full chain structures (1D6G, 1ORS, 2AHY, 1R3J, 1S5H), very low-resolution structures (>4.2 Å, 1IFK), miss-folded structures (1ORQ and 2A01) and proteins with membrane border boundary errors (2QFI) or problems in geometry (1YEW).

Uniprot sequences corresponding to the remaining sequences were used to search for homologs by running three rounds of PSI-Blast[79] using a conservative E-value cutoff of 10^{-5} against uniref90[80] from November 2007. Chains with less than four identified homologs, mostly consisting of short transmembrane proteins, were removed in a second filtering step. Thereafter, the structure of highest resolution from each superfamily was chosen as representatives. Finally, cd-hit[81] was used for homology reduction at 40% sequence identity, leaving 85 chains from 59 structures.

General analysis

As in our previous studies,[4–6] all proteins were oriented so that the predicted membrane core was located at the X-Y plane, thus the proteins could easily be studied as a function of the Z-coordinate. Using the definition from our study of deep core coils,[6] the OPM method, which is based on energetic calculations, was

used to define the orientation.[78] The *Z*-coordinates were used to classify all residues into three main groups: globular (Z > 22Å), lipid–water interface (10 Å < Z < 22 Å), and membrane core (Z < 10Å). The final dataset contained 21,264 residues, with 5,487 in the core, 6,945 in the lipid–water interface, and 8,832 outside the membrane region. The residues were grouped by hydrophobicity[15] into hydrophobic [A, C, F, I, L, M and V], weakly polar [G, Y, W, S and T] and strongly polar [D, E, K, R, H, N, P and Q]. This resulted in 545 strongly polar residues in the membrane core.

The functional classification of the proteins in this study was mainly based on the enzyme (EC) and transporter (TC) classification systems. The information was extracted from PDB[82] and TCDB,[83] respectively. As the majority of enzymes were oxidoreductases involved in respiration, they were grouped as a separate class in order not to bias the enzyme group. Further, photosynthesis related–proteins were grouped together as they have many common features in their structure and function. Remaining proteins were assigned as receptors or others based on PDBsum.[84]

The amino acid substitution rates were estimated as described in,[6] by using the PSI-BLAST derived multiple sequence alignments (mentioned above) as input to rate4site.[85] The residue values from rate4site were normalized for each protein.

Surface accessibility was calculated by Naccess 2.1.1[86] with probe size 1.4 Å, corresponding to the size of water, over the whole structure. The relative accessible surface area (RSA) was obtained directly from Naccess, where the accessible surface area of a residue is normalized by an extended Ala-X-Ala tri-peptide conformation. The mean relative surface accessibility was then 30.8, with 24.9 in the core, 31.6 in the lipid–water interface, and 34.1 outside the membrane. All residues exposed to less than 25% were referred to as buried and all others as exposed.

Atom–atom distances for all pairs within 6 Å were calculated using the program contact from the CCP4 6.0 package[87] with following settings: mode ALL, limits 0 and 6 Å. For each polar atom in the sidechain of standard residues, the closest polar atom within 3.6 Å was selected. No angle restraints were used. All atoms except C, H, and S were defined as polar atoms. However, here Cys residues were considered as polar, if they do not participate in disulphide bonds. Polar atoms without an interaction partner were classified as putative solvent.

Functional annotations

A "key reference" article was selected for each of the 59 protein structures according to recommendations by PDBsum.[84] For several proteins, the literature study was extended by including articles of earlier structures and recent reviews. Functional importance was assigned to residues that were depicted by authors to be directly involved in the function or to the ones that bind key part of prosthetic groups. Residues that "only" stabilize structure of active site or form polar inner surfaces were not assigned as directly functional.

Assignment of local environment

As described above, all residues exposed to less than 25% were referred to as buried and all others as exposed. In addition to that the local environment around the polar residues in the membrane core were assigned based on visual inspection of the structures. The polar groups were defined as cavity exposed, if at least one polar atom of the residue was in contact with at least one water or at least one polar atom is located next to the cavity. Similarly, polar residues were defined to not have proper bonding partners, if at least one polar atom was not having polar contact nor lining a water filled cavity. For prolines unpaired mainchain carbonyls were defined as its "polar groups."

Predictions in human membrane proteins

The sequences and functional classification of all human membrane proteins were obtained from Almen et al. (2009).[88] Homolog identification were performed by blast and conservation analysis with rate4site with same parameters as given for the OPM-dataset. Zpred[24] was used to predict the distance from membrane core and the number of TMs was estimated using octopus.[89] Accessibility was predicted by MPRAP.[26] For calculation of accuracy and MCC values the following assignments were used: true positives have *Z*-coordinate <= 10 and Zpred <= 10, false positives have Z-coordinate > 10 and Zpred <= 10, true negatives have Z-coordinate > 10 and Zpred > 10 and false negatives have Z-coordinate <= 10 and Zpred > 10. *P*-values for over-representation of polar rich proteins were calculated using Fisher's exact test in *R*.[90]

Data analysis and visualization

The molecular illustrations were created with PyMol (DeLano, W.L. The PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, USA). The remaining figures as well as the statistical tests were generated in R.[90]

ACKNOWLEDGMENTS

The authors thank Sara Light for useful suggestions regarding the manuscript.

REFERENCES

1. Granseth E, Daley D, Rapp M, Melen K, von Heijne G. Experimentally constrained topology models for 51,208 bacterial inner membrane proteins. J Mol Biol 2005;352:489–494.
2. Elofsson A, von Heijne G. Membrane protein structure: prediction versus reality. Annu Rev Biochem 2007;76:125–140.

3. Yohannan S, Faham S, Yang D, Whitelegge J, Bowie J. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. Proc Natl Acad Sci USA 2004; 101:959–963.
4. Granseth E, von Heijne G, Elofsson A. A study of the membrane-water interface region of membrane proteins. J Mol Biol 2005;346:377–385.
5. Viklund H, Granseth E, Elofsson A. Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. J Mol Biol 2006;361:591–603.
6. Kauko A, Illergard K, Elofsson A. Coils in the membrane core are conserved and functionally important. J Mol Biol 2008;380:170–180.
7. Honig B, Yang A. Free energy balance in protein folding. Adv Protein Chem 1995;46:27–58.
8. Lins L, Brasseur R. The hydrophobic effect in protein folding. FASEB J 1995;9:535–540.
9. Stevens T, Arkin I. Are membrane proteins "inside-out" proteins? Proteins 1999;36:135–143.
10. Wallin E, Tsukihara T, Yoshikawa S, von Heijne G, Elofsson A. Architecture of helix bundle membrane proteins. An analysis of cytochrome c oxidase from bovine mitochondria. Protein Sci 1997;6:808–815.
11. White S, Wimley W. Membrane protein folding and stability: physical principles. Annu Rev Biophys Biomol Struct 1999;28:319–365.
12. 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.
13. Isom D, Cannon B, Castaneda C, Robinson A, Garcia-Moreno B. High tolerance for ionizable residues in the hydrophobic interior of proteins. Proc Natl Acad Sci USA 2008;105:17784–17788.
14. White S, Wimley W. Hydrophobic interactions of peptides with membrane interfaces. Biochim Biophys Acta 1998;1376:339–352.
15. Hessa T, Meindl-Beinker N, Bernsel A, Kim H, Sato Y, Lerch-Bader M, Nilsson I, White S, von Heijne G. Molecular code for transmembrane-helix recognition by the Sec61 translocon. Nature 2007;450:1026–1030.
16. Oberai A, Joh N, Pettit F, Bowie J. Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. Proc Natl Acad Sci USA 2009;106:17747–17750.
17. Bernsel A, Viklund H, Falk J, Lindahl E, von Heijne G, Elofsson A. Prediction of membrane-protein topology from first principles. Proc Natl Acad Sci USA 2008;105:7177–7181.
18. Jones D, Taylor W, Thornton J. A mutation data matrix for transmembrane proteins. FEBS Lett 1994;339:269–275.
19. Partridge A, Therien A, Deber C. Missense mutations in transmembrane domains of proteins: phenotypic propensity of polar residues for human disease. Proteins 2004;54:648–656.
20. Adamian L, Nanda V, DeGrado W, Liang J. Empirical lipid propensities of amino acid residues in multispan alpha helical membrane proteins. Proteins 2005;15:496–509.
21. Johansson A, Lindahl E. Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. Biophys J 2006;91:4450–4463.
22. Vijayakumar M, Qian H, Zhou H. Hydrogen bonds between short polar side chains and peptide backbone: prevalence in proteins and effects on helix-forming propensities. Proteins 1999;34:497–507.
23. Presta LG, Rose GD. Helix signals in proteins. Science 1988;240:1632–1641.
24. Granseth E, Viklund H, Elofsson A. ZPRED: predicting the distance to the membrane center for residues in alpha-helical membrane proteins. Bioinformatics 2006;22:e191–e196.
25. Papaloukas C, Granseth E, Viklund H, Elofsson A. Estimating the length of transmembrane helices using Z-coordinate predictions. Protein Sci 2008;17:271–278.
26. Illergard K, Callegari S, Elofsson A. MPRAP: an accessibility predictor for alpha-helical transmembrane proteins that performs well inside and outside the membrane. BMC Bioinformatics 2010;11:333.
27. Sandelin E. On hydrophobicity and conformational specificity in proteins. Biophys J 2004;86(1 Part 1):23–30.
28. Rosenbaum D, Rasmussen S, Kobilka B. The structure and function of G-protein-coupled receptors. Nature 2009;459:356–363.
29. Topiol S, Sabio M. X-ray structure breakthroughs in the GPCR transmembrane region. Biochem Pharmacol 2009;78:11–20.
30. Oldham M, Davidson A, Chen J. Structural insights into ABC transporter mechanism. Curr Opin Struct Biol 2008;18:726–733.
31. Apell H. How do P-type ATPases transport ions? Bioelectrochemistry 2004;63:149–156.
32. Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trezeguet V, Lauquin G, Brandolin G. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. Nature 2003; 426:39–44.
33. Boudker O, Ryan R, Yernool D, Shimamoto K, Gouaux E. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. Nature 2007;445:387–393.
34. Toyoshima C, Nomura H, Tsuda T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. Nature 2004;432:361–368.
35. Yamashita A, Singh S, Kawate T, Jin Y, Gouaux E. Crystal structure of a bacterial homologue of Na+/Cl–dependent neurotransmitter transporters. Nature 2005;437:215–223.
36. Abramson J, Smirnova I, Kasho V, Verner G, Kaback H, Iwata S. Structure and mechanism of the lactose permease of Escherichia coli. Science 2003;301:610–615.
37. Seeger M, Schiefner A, Eicher T, Verrey F, Diederichs K, Pos K. Structural asymmetry of AcrB trimer suggests a peristaltic pump mechanism. Science 2006;313:1295–1298.
38. Jormakka M, Tornroth S, Byrne B, Iwata S. Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. Science 2002;295:1863–1868.
39. Huang L, Cobessi D, Tung E, Berry E. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc1 complex: a new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. J Mol Biol 2005;351:573–597.
40. Bertero M, Rothery R, Palak M, Hou C, Lim D, Blasco F, Weiner J, Strynadka N. Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A. Nat Struct Biol 2003;10:681–687.
41. Tsukihara T, Aoyama H, Yamashita E, Tomizaki T, Yamaguchi H, Shinzawa-Itoh K, Nakashima R, Yaono R, Yoshikawa S. The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 A. Science 1996;272:1136–1144.
42. Shimokata K, Katayama Y, Murayama H, Suematsu M, Tsukihara T, Muramoto K, Aoyama H, Yoshikawa S, Shimada H. The proton pumping pathway of bovine heart cytochrome c oxidase. Proc Natl Acad Sci USA 2007;104:4200–4205.
43. Huang L, Sun G, Cobessi D, Wang A, Shen J, Tung E, Anderson V, Berry E. 3-nitropropionic acid is a suicide inhibitor of mitochondrial respiration that, upon oxidation by complex II, forms a covalent adduct with a catalytic base arginine in the active site of the enzyme. J Biol Chem 2006;281:5965–5972.
44. Lancaster C, Kroger A, Auer M, Michel H. Structure of fumarate reductase from Wolinella succinogenes at 2.2 A resolution. Nature 1999;402:377–385.
45. Madej M, Nasiri H, Hilgendorff N, Schwalbe H, Lancaster C. Evidence for transmembrane proton transfer in a dihaem-containing membrane protein complex. EMBO J 2006;25:4963–4970.
46. Holm P, Bhakat P, Jegerschold C, Gyobu N, Mitsuoka K, Fujiyoshi Y, Morgenstern R, Hebert H. Structural basis for detoxification and oxidative stress protection in membranes. J Mol Biol 2006;360:934–945.
47. Wang Y, Zhang Y, Ha Y. Crystal structure of a rhomboid family intramembrane protease. Nature 2006;444:179–180.
48. Lemieux M, Fischer S, Cherney M, Bateman K, James M. The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. Proc Natl Acad Sci USA 2007;104:750–754.

49. Feng L, Yan H, Wu Z, Yan N, Wang Z, Jeffrey P, Shi Y. Structure of a site-2 protease family intramembrane metalloprotease. Science 2007;318:1608–1612.
50. Jordan P, Fromme P, Witt H, Klukas O, Saenger W, Krauss N. Three-dimensional structure of cyanobacterial photosystem I at 2.5 A resolution. Nature 2001;411:909–917.
51. Prince S, Papiz M, Freer A, McDermott G, Hawthornthwaite-Lawless A, Cogdell R, Isaacs N. Apoprotein structure in the LH2 complex from *Rhodopseudomonas acidophila* strain 10050: modular assembly and protein pigment interactions. J Mol Biol 1997;268:412–423.
52. Komiya H, Yeates T, Rees D, Allen J, Feher G. Structure of the reaction center from *Rhodobacter sphaeroides* R-26 and 2.4.1: symmetry relations and sequence comparisons between different species. Proc Natl Acad Sci USA 1988;85:9012–9016.
53. Loll B, Kern J, Saenger W, Zouni A, Biesiadka J. Towards complete cofactor arrangement in the 3.0 A resolution structure of photosystem II. Nature 2005;438:1040–1044.
54. Standfuss J, Terwisscha van Scheltinga A, Lamborghini M, Kuhlbrandt W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 A resolution. EMBO J 2005;24:919–928.
55. Luecke H, Schobert B, Richter H, Cartailler J, Lanyi J. Structure of bacteriorhodopsin at 1.55 A resolution. J Mol Biol 1999;291:899–911.
56. Van den Berg B, Clemons W, Jr, Collinson I, Modis Y, Hartmann E, Harrison S, Rapoport T. X-ray structure of a protein-conducting channel. Nature 2004;427:36–44.
57. Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. Nature 2000;405:647–655.
58. Murata T, Yamato I, Kakinuma Y, Leslie A, Walker J. Structure of the rotor of the V-Type Na+-ATPase from *Enterococcus hirae*. Science 2005;308:654–659.
59. Hollenstein K, Frei D, Locher K. Structure of an ABC transporter in complex with its binding protein. Nature 2007;446:213–216.
60. Oldham M, Khare D, Quiocho F, Davidson A, Chen J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature 2007;450:515–521.
61. Huang Y, Lemieux M, Song J, Auer M, Wang D. Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli*. Science 2003;301:616–620.
62. Law C, Almqvist J, Bernstein A, Goetz R, Huang Y, Soudant C, Laaksonen A, Hovmoller S, Wang D. Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. J Mol Biol 2008;378:828–839.
63. D'Rozario R, Sansom M. Helix dynamics in a membrane transport protein: comparative simulations of the glycerol-3-phosphate transporter and its constituent helices. Mol Membr Biol 2008;25:571–583.
64. Hunte C, Screpanti E, Venturi M, Rimon A, Padan E, Michel H. Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature 2005;435:1197–1202.
65. Yin Y, He X, Szewczyk P, Nguyen T, Chang G. Structure of the multidrug transporter EmrD from *Escherichia coli*. Science 2006;312:741–744.
66. Dutzler R, Campbell E, MacKinnon R. Gating the selectivity filter in ClC chloride channels. Science 2003;300:108–112.
67. Lee J, Kozono D, Remis J, Kitagawa Y, Agre P, Stroud R. Structural basis for conductance by the archaeal aquaporin AqpM at 1.68 A. Proc Natl Acad Sci USA 2005;102:18932–18937.
68. Tajkhorshid E, Nollert P, Jensen M, Miercke L, O'Connell J, Stroud R, Schulten K. Control of the selectivity of the aquaporin water channel family by global orientational tuning. Science 2002;296:525–530.
69. Jasti J, Furukawa H, Gonzales E, Gouaux E. Structure of acid-sensing ion channel 1 at 1.9 A resolution and low pH. Nature 2007;449:316–323.
70. Paukert M, Babini E, Pusch M, Grunder S. Identification of the Ca2+ blocking site of acid-sensing ion channel (ASIC) 1: implications for channel gating. J Gen Physiol 2004;124:383–394.
71. Long S, Tao X, Campbell E, MacKinnon R. Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment. Nature 2007;450:376–382.
72. Lunin V, Dobrovetsky E, Khutoreskaya G, Zhang R, Joachimiak A, Doyle D, Bochkarev A, Maguire M, Edwards A, Koth C. Crystal structure of the CorA Mg2+ transporter. Nature 2006;440:833–837.
73. Hattori M, Tanaka Y, Fukai S, Ishitani R, Nureki O. Crystal structure of the MgtE Mg2+ transporter. Nature 2007;448:1072–1075.
74. Cherezov V, Rosenbaum D, Hanson M, Rasmussen S, Thian F, Kobilka T, Choi H, Kuhn P, Weis W, Kobilka B, Stevens R. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science 2007;318:1258–1265.
75. Call M, Schnell J, Xu C, Lutz R, Chou J, Wucherpfennig K. The structure of the zetazeta transmembrane dimer reveals features essential for its assembly with the T cell receptor. Cell 2006;127:355–368.
76. Dessailly B, Lensink M, Wodak S. Relating destabilizing regions to known functional sites in proteins. BMC Bioinformatics 2007;8:141.
77. Cheng G, Qian B, Samudrala R, Baker D. Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design. Nucleic Acids Res 2005;33:5861–5867.
78. Lomize M, Lomize A, Pogozheva I, Mosberg H. OPM: orientations of proteins in membranes database. Bioinformatics 2006;22:623–625.
79. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
80. Suzek B, Huang H, McGarvey P, Mazumder R, Wu C. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 2007;23:1282–1288.
81. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658–1659.
82. Sussman J, Lin D, Jiang J, Manning N, Prilusky J, Ritter O, Abola E. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallogr D Biol Crystallogr 1998;54(Part 6 Part 1):1078–1084.
83. Saier M, Jr, Tran C, Barabote R. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res 2006;34:D181–6.
84. Laskowski R, Chistyakov V, Thornton J. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Res 2005;33:D266–8.
85. Mayrose I, Graur D, Ben-Tal N, Pupko T. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. Mol Biol Evol 2004;21:1781–1791.
86. Hubbard SJ, T. J. NACCESS, Computer program. Department of Biochemistry and Molecular Biology. Available at:http://wolf.bi.umist.ac.uk/unix/naccess.html. Accessed on: 1–2, Jan 1993.
87. Collaborative Computational Project N.. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50 (Part 5):760–763.
88. Almen M, Nordstrom K, Fredriksson R, Schioth H. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol 2009;7:50.
89. Viklund H, Elofsson A. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 2008;24:1662–1668.
90. R Development Core Team 2006. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2005.