Journal of proteome research

TUTORIAL

pubs.acs.org/jpr

An Introduction to Membrane Proteins[†]

Linnea E. Hedin, Kristoffer Illergård, and Arne Elofsson*

Department of Biochemitry and Biophysics, Stockholm Bioinformatics Center, Center for Biomembrane Research, Science for life laboratory, Swedish E-science Research Center, Stockholm University, 106 91 Stockholm, Sweden

Supporting Information

ABSTRACT: α-Helical membrane proteins are important for many biological functions. Due to physicochemical constraints, the structures of membrane proteins differ from the structure of soluble proteins. Historically, membrane protein structures were assumed to be more or less two-dimensional, consisting of long, straight, membrane-spanning parallel helices packed against each other. However, during the past decade, a number of the new membrane protein structures cast doubt on this notion. Today, it is evident that the structures of many membrane proteins are equally complex as for many soluble proteins. Here, we review this development and discuss the consequences for our understanding of membrane protein biogenesis, folding, evolution, and bioinformatics.

KEYWORDS: membrane proteins, bioinformatics, protein structure, protein evolution, translocon insertion, protein biogenesis

INTRODUCTION

Biological membranes function as barriers surrounding living cells and organelles in eukaryotic cells. The basic unit of most biological membranes is the phospholipids, but they also contain a large fraction of proteins that are interspersed among or loosely attached to the phospholipids. Actually, proteins embedded within the lipids can make up half of the mass of a biological membrane.[1] The environment surrounding the lipid-embedded proteins is obviously very different from the environment around water-soluble proteins. Soluble proteins are found in a watery, hydrophilic, environment, while membrane proteins are found within a lipid, hydrophobic, bilayer. The differences in environment have fundamental effects on the structure of the proteins. Membrane proteins can be classified as peripheral or integral. Peripheral membrane proteins are attached to the membrane either loosely through electrostatic or van der Waals interactions with the lipid head-groups or other membrane proteins, or through a covalent anchor. Integral membrane proteins span the membrane.[1] Two types of integral membrane proteins have been identified, α-helical and β-barrels. This review will mainly discuss α-helical membrane proteins.

α-Helical membrane proteins are the most abundant and also the best studied. They contain one or several α-helices. These helices are hydrophobic and typically about 20 residues long. These characteristics make them relatively easy to detect from sequence. Using prediction methods based on hydrophobicity, it can be estimated that 20−30% of the genes in most organisms encode α-helical membrane proteins.[2,3]

β-Barrel proteins consist of β-sheets forming a barrel-like structure. Here, the residues in each sheet point alternatively outward, facing the lipids, and inward, facing the inside of the barrel, resulting in a sequence pattern in which the residues are typically alternatively hydrophobic and polar. The outcome is a polar channel through which water-soluble molecules can cross. It has been estimated that around 2−3% of the genes in gram-negative bacteria encodes β-barrel membrane proteins.[4−6] Further, a handful of β-barrel proteins can also be found in the outer membrane of mitochondria and chloroplasts.

Integral membrane proteins are important for all cellular life, performing many crucial functions, Figure 1. They transport ions, metabolites, and larger molecules such as proteins and RNA across the membranes. Membrane proteins are also responsible for sending and receiving chemical signals, propagating electrical impulses, attaching cells to each other, and anchoring other proteins to specific locations in the cell. Other functions include regulating intracellular vesicular transport, controlling membrane lipid composition, and organizing and maintaining the shape of organelles and the cell itself.[1] It should also be noted that membrane proteins are harder to work with experimentally than soluble proteins, due to their hydrophobic nature. This has resulted in an under-representation of membrane proteins among proteins of known structure.

BIOLOGICAL MEMBRANES

Membranes need to be mechanistically strong and flexible and impermeable to unwanted compounds. Further, to enable transport of compounds and signals there is also a need for a leakage mechanism.[1] The basic components of biological membranes are lipid molecules, consisting of a hydrophilic headgroup, and one or more hydrophobic fatty acid hydrocarbon tails. As a

[†] This Tutorial is part of the International Proteomics Tutorial Program (IPTP7).

Published: August 05, 2011

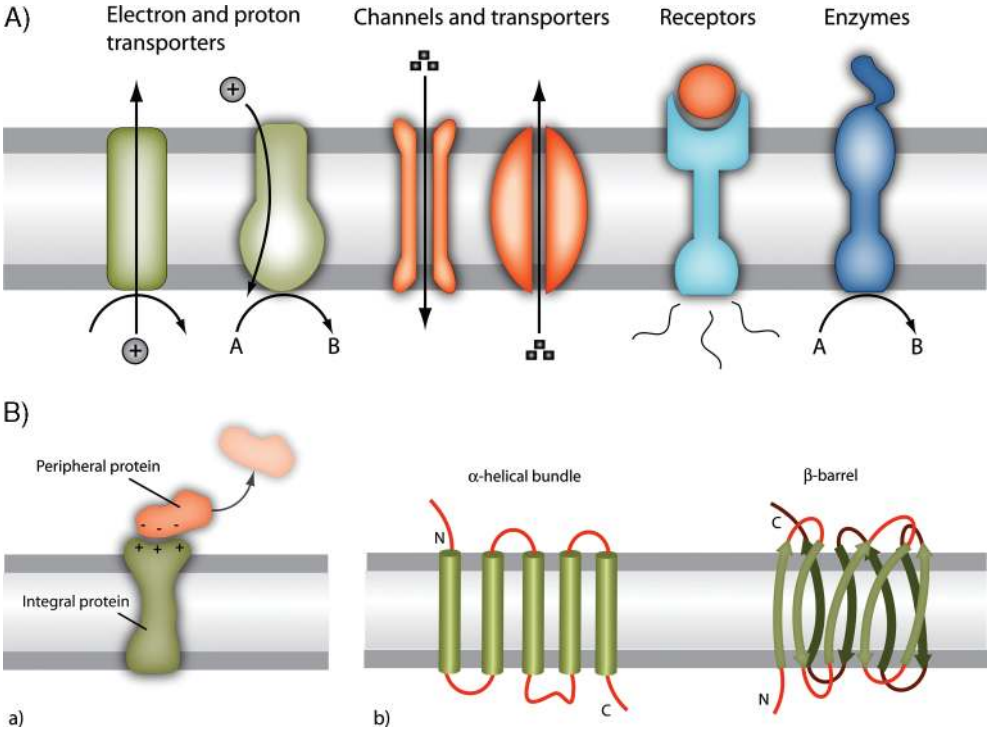

Figure 1. (A) Functions of membrane proteins are diverse. The proteins are, for example, involved in energy transport during photosynthesis or respiration, transport of molecules across their membranes, transmission of chemical signals or catalyze of chemical reactions. Their important functions make them highly interesting from a medical point of view. (B) Membrane proteins can be integral or peripheral. Integral membrane proteins come in two flavors: α-helical bundles and β-barrels.

consequence of their amphiphilic nature, lipids aggregate and spontaneously self-organize in water, generally into either micelles or bilayers. In cells, the hydrophobic effect will cause them to form bilayers, consisting of two sheets of lipids with their hydrophobic tails facing each other.[1] The hydrophilic headgroups shield the tails, forming an interface between the core of the membrane and the aqueous surroundings.[7] There is no sharp border between the hydrophobic core and the surrounding water as this interface region provides a zone of gradually changing hydrophobicity. In 1992, Wiener and White determined the structure of a bilayer of pure DOPC lipids,[8] Figure 2. The hydrophobic core was about 30 Å thick, while the interface region extended about 15 Å on either side. The tails of the lipids can be of different lengths and stiffness and the size and electrical charge of the headgroup can differ. Steroidic lipids, like cholesterol, are very rigid and confer stability. The features and properties of a biological membrane—thickness, fluidity, curvature, pressure—will vary with lipid composition.[9] It varies from organism to organism and between organelles in the same cell, even between leaflets of the same bilayer. It is worth mentioning that Archaea, often found in extreme habitats, employ lipids that differ substantially from the bacterial and eukaryotic ones, even though the hydrophilic part of the headgroups can be the same.[10]

■ BIOGENESIS OF α-HELICAL MEMBRANE PROTEINS

The biogenesis of membrane proteins differs from the biogenesis of soluble proteins. The vast majority of membrane proteins are synthesized in the cytosol, where the ribosomes translate mRNA codons into amino acids and add them to the growing polypeptide chain. In general, soluble cytosolic proteins simply fold as they emerge from the ribosome. For proteins meant for secretion or integration, the process is more complicated and there are both cotranslational and post-translational pathways for translocation across or insertion into the membrane.

The major system for membrane integration into the plasma membrane is the SRP-dependent Sec pathway. Its major components are the signal recognition particle (SRP), the signal recognition particle receptor (SR) and the secretase (Sec) translocon, see Figure 3. After translation is initiated, two major steps follows:

1. Targeting to the translocon. The SRP will recognize and bind a highly hydrophobic stretch of the nascent polypeptide chain, the signal peptide, when it emerges from the ribosome. In eukaryotes, but not in prokaryotes, this arrests the chain elongation until the ribosome contacts the translocon. SRP then binds to its receptor in the ER membrane (eukaryotes) or in the plasma membrane (prokaryotes), which catalyzes the transfer of the ribosome-nascent chain-complex to the translocon.
2. Translocation/integration. After transfer to the translocon, the exit tunnel in the ribosome is aligned with the translocon pore and the nascent molecule is threaded directly through the channel as it is being synthesized.[11] Since protein translation and translocation or membrane insertion occur simultaneously, the process is termed cotranslational.[1] When the signal peptide reaches the translocon it either reorients, placing the N-terminal in the cytosol, giving the protein an N_{in}-orientation, or it does not, resulting in an N_{out}-orientation. As translation and chain elongation proceed, the nascent chain passes through the translocon channel. Segments that are of sufficient length and hydrophobicity will exit the channel laterally and enter the lipid bilayer.[11]

dx.doi.org/10.1021/pr200145a |J. Proteome Res. 2011, 10, 3324–3331

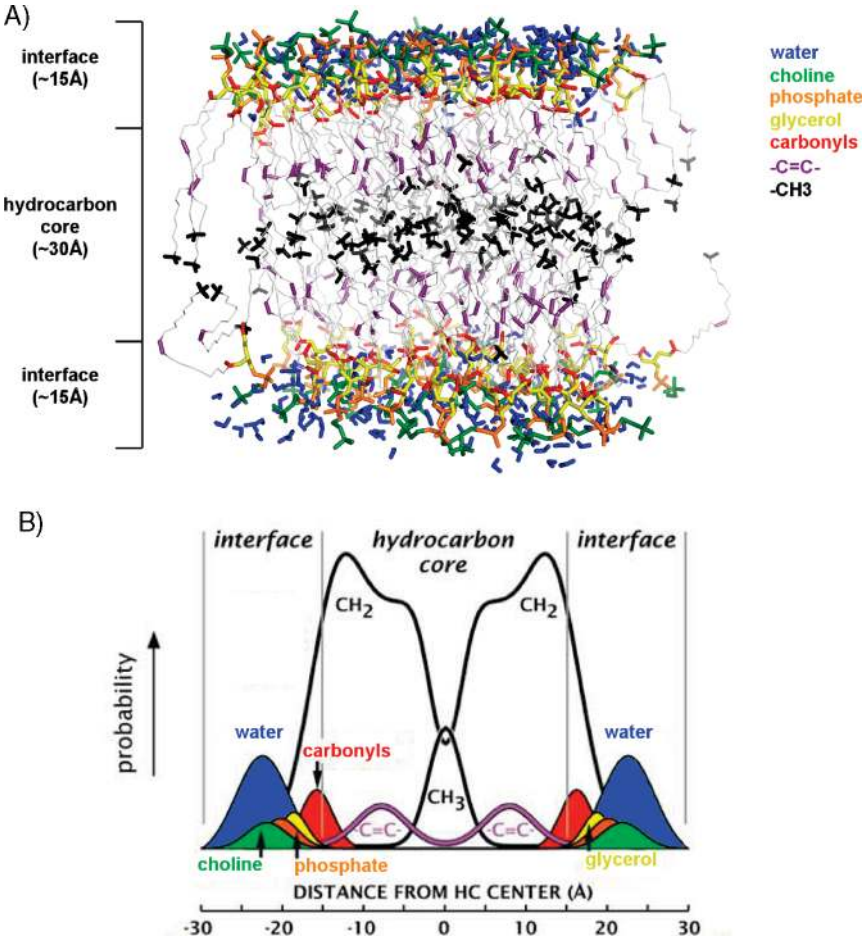

Figure 2. (A) Structure of 1,2-Dioleoyl-sn-Glycero-3-phosphocholine (DOPC) lipid bilayer after 1.5 ns molecular dynamics simulation. The figure is adapted from Feller, 1997. (B) Distribution of the functional groups across a DOPC lipid bilayer, as determined by X-ray and neutron diffraction measurements. The colors in both figures should represent similar groups. The figure is adapted from Wiener 1992.

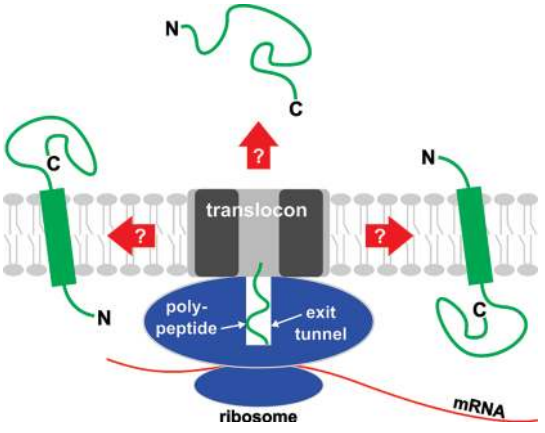

Figure 3. Translocon apparatus can be seen as a switching station that receives elongating peptide sequences from the ribosome, and directs them either across the membrane or into the membrane bilayer, depending upon the segment's properties. The figure is influenced by White, 1999.

It has been assumed that transmembrane helices move into the lipid membrane from the protein-conducting channel through a partitioning process, where sufficiently hydrophobic helices prefer the bilayer, whereas more polar helices favor the translocon. Therefore, the process would appear as an equilibrium process in which favorable interactions between lipids and amino acid side chains promote membrane integration rather than translocation.[11] The efficiency with which a transmembrane segment is integrated by the translocon should then be sensitive to the relative positions of amino acids within the segment. This trend is evident when looking at amino acid distributions in known membrane protein structures.[12]

For single-spanning proteins, it seems that the hydrophobicity, according to the biological hydrophobicity scale, determines whether or not a given polypeptide segment will form a transmembrane segment. For multispanning proteins, the known three-dimensional structures show that a surprisingly large fraction of their membrane-spanning segments do not follow this trend. Indeed, more than 25 segments in multispanning proteins may depend on interactions with other parts of the same peptide chain for efficient insertion.[11] A few examples where the interaction with neighboring helices aids the insertion have been found.[13]

Other Systems for Insertion of Membrane Proteins

Even though most membrane proteins seem to be inserted via the SRP-dependent Sec pathway, there are other insertion mechanisms. These are found in the bacterial membranes and in the membranes of the eukaryotic mitochondria, chloroplasts

dx.doi.org/10.1021/pr200145a |J. Proteome Res. 2011, 10, 3324–3331

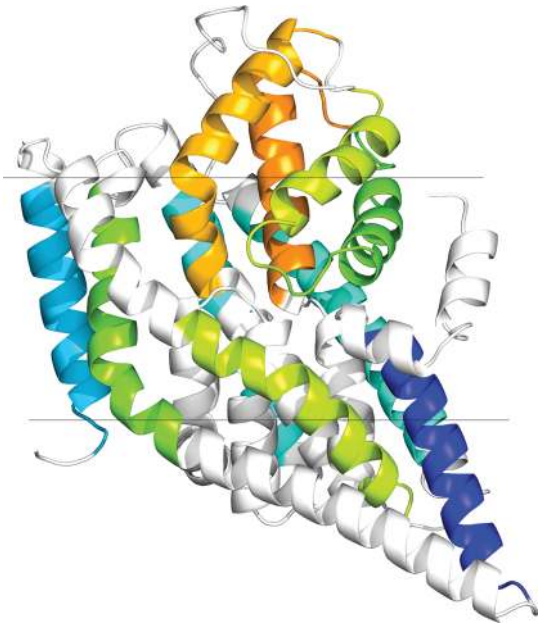

Figure 4. Structure of the Glutamate transporter homologue. The colored segments correspond to the most hydrophobic regions of each transmembrane helix. The more bluish the more hydrophobic is the segment. The location of the membrane borders are indicated with gray lines. During the positioning of transmembrane regions in the folded glutamate homologue, structure does clearly not always correspond to hydrophobic region.

and peroxisomes. A very brief review of these systems will be given here.

Mitochondria lack the Sec translocon completely and instead this dual-membrane organelle is equipped with an elaborate system for import and insertion of membrane proteins, involving six major protein complexes and a number of chaperons.[14]

In peroxisomes, some membrane proteins are imported cotranslationally, in a not yet completely elucidated process, while others come in vesicles budded off from the ER.[15]

Chloroplasts have translocons of the inner and outer membranes (TOC and TIC) as well as the twin arginine translocon (Tat) that translocates folded proteins over the thylakoid membrane and seems to be involved in the insertion of some integral membrane proteins. The Tat system is also found in Eubacteria, Archaea, and plant mitochondria.[16]

Folding of Membrane Proteins

The selection of transmembrane segments is only the first step in the complex process of bringing the transmembrane segments together to form the native protein structure. After translocon-guided insertion, the transmembrane segments of multispanning membrane proteins must condense to form the final folded three-dimensional structure. Originally, it was believed that the folding process of membrane proteins was quite simple. However, recent studies have shown that large rearrangements are quite common.[17] Also, it has been indicated experimentally that some helices are not inserted cotranslationally but only later during the folding process.[13,18] It is quite likely that reentrant regions[19] also are inserted into the membrane post-translationally. One particular intriguing protein is the Glutamate transporter homologue, see Figure 4. Here, residues 151−171 are recognized as the fourth helix by the translocon. Later, during folding, this helix is shifted so that residues 129−160 constitute the transmembrane helix.[17] It has also been found that Aquaporin 1 undergoes large structural rearrangements after translocon recognition.[13,18] Here, the second helix is not recognized by the translocon and consequently helix number 3 is inserted in the wrong orientation. Later on, helix 2 inserts and helix 3 flips. The molecular details of this process are not known.

Stabilizing Forces of Membrane Proteins

During the folding of a protein some residues will remain exposed to the environment while others will become buried in the protein interior. For water-soluble proteins it is energetically favorable to bury hydrophobic residues and expose polar and charged residues surrounding water. However, transmembrane proteins face three distinct environments: a hydrophobic lipid environment inside the membrane, a hydrophilic water environment outside the membrane, and an interface region rich in phospholipid headgroups. As a result of these distinct environments, the surfaces of transmembrane proteins need to expose different types of residues at different locations. However, the interior of transmembrane proteins could, at least for stability, remain identical to what is found in the interior of soluble proteins.

Early studies of the bacteriorhodopsin structure suggested that membrane proteins are "inside-out", that is, that they consist of a hydrophilic interior and a hydrophobic exterior. However, later studies indicate that the "inside-out" rule is not generally applicable to all α-helical membrane proteins.[20−22] The dominant driving force behind folding of water-soluble proteins is the hydrophobic effect that minimizes unfavorable interactions between hydrophobic residues and water.[23] In the membrane environment, this driving force does not exist, nor is there any major driving force to bury polar residues within the protein interior, as the solvent environment there does not differ significantly from the one of the membrane core.

In the absence of a strong hydrophobic effect, it has been proposed that hydrogen bonds or ionic interactions are the main driving forces in membrane protein folding. However, even if hydrogen bonds certainly are important, recent studies have indicated that hydrogen bonds are not significantly stronger within the membrane than outside.[24] This is in contrast to what has been observed in apolar solvents, indicating that such solvents do not accurately approximate the membrane environment. This might also explain the observed difference between the biological hydrophobicity scale[25,26] and other hydrophobicity scales, most of which were derived from experimental setups utilizing apolar solvents. In general, both experimental and theoretical studies in these solvents indicate a significantly higher energetic cost for inserting polar groups into a lipid environment than what the biological hydrophobicity scale predicts. However, when additional proteins are included in the description of the lipid environment, this discrepancy disappears.[27]

The existence of a number of motifs involving small side chains has been proposed to be important for membrane protein stability.[28] These motifs allow tight packing of TM helices. Short chained-residues are spaced in such a way that they end up on the same side of their helix, that is, in positions i and $i+4$ or in i and $i+7$, creating a surface that allows close proximity. Statistics show that Glycine zippers occur in more than 10% of all known MP structures.[29] These motifs usually form right-handed crossings of the helices. Tightly packed helices are stabilized by stronger van der Waals interactions due to closer proximity of the interact-

ing helices. Many helix−helix interfaces are also stabilized by hydrogen bonds, both "classical" bonds and the weaker bonds involving the peptide backbone, Cα−H−O.[30] It has been shown that the residues participating in hydrogen bonds are more conserved than other buried residues and that the packing preferences differ between channels and transporters and other membrane proteins, mostly because these classes of protein are enriched in different amino acids.[31,32] In channels and transporters, the helix−helix interfaces are not as tightly packed, hydrogen bonds occur more frequently and are often found in proximity to water-filled cavities, providing alternative bonding partners and allowing weaker bonds suitable for the structural rearrangements needed for functionality.[33]

STRUCTURE OF MEMBRANE PROTEINS

For a long time the general view was that α-helical membrane proteins form simple helix bundles, with their hydrophobic transmembrane helices crossing the membrane in more or less perpendicular orientations. Indeed, many membrane proteins, abide by this principle. Bacteriorhodopsin is one example of such a protein. It is a seven-helix bundle and functions as a light-driven proton pump: small, light-induced movements in its transmembrane helices entice protons to translocate across the membrane against an electrochemical gradient. The helices lie almost straight in the membrane and pack with typical knobs-into-holes packing angles.[11]

However, some more recently solved membrane protein structures show that the structural repertoire is much more complex. The first such example was discovered when the structure of aquaporin was solved in 2000.[34] In addition to the six long-standard transmembrane helices, aquaporin contains two "reentrant" regions. Such reentrant regions span only part of the membrane and enter and exits the lipid bilayer on the same side. Later studies have shown that membrane-embedded helices can be short, long, kinked or interrupted in the middle of the membrane (coils), almost perpendicular to the membrane plane, strongly tilted, laying flat on the surface of the membrane (interfacial helices), or even span only a part of the membrane and then turn back as in the case of aquaporin. In addition, some helices are packed in manners that do not follow the simple "knobs-into-holes" geometry.[6]

The glutamate transporter homologue is one protein with many such deviations and, as Figure 4 shows, it has a complex topology. This protein contains long, steeply inclined helices and short, closely spaced pairs of helices forming reentrant regions. There are also stretches of nonhelical structures, deep within the membrane that are largely buried between the transmembrane helices.[6] In summary, structural irregularities include kinked TM-helices,[35] interfacial helices,[36] reentrant regions,[19] and coils in the membrane.[33]

Amino Acid Preferences of α-Helical Membrane Proteins

To fold properly, amino acid frequencies in membrane proteins have adopted to the three different environments surrounding them. This has resulted in that different amino acids are preferred at different locations along the normal of the membrane surface. In short, hydrophobic amino-acids are frequent in the core of the membrane, while tyrosine and tryptophan are frequent in the polar headgroup (interface) regions and amino acid distributions outside the membrane resembles the one of soluble proteins, see Figure 5. Charged residues also have statistically preferred locations, and we will discuss the so-called snorkeling effect, the positive inside rule and the importance of these residues for function.

Even if membrane proteins are very hydrophobic within the membrane core region and about 70% of all amino acids in the core of the membrane are hydrophobic, see Figure 5, exceptions to this and the other statistical patterns mentioned above are seen as more and more three-dimensional structures are solved. We again want to remind the reader that charged residues and sequences causing irregular secondary structure within the hydrophobic core have been observed.[33,37]

Tyrosine and Tryptophan are more abundant about 15−20 Å away from the center of the membrane, that is, within the interface region. This rather tight anchoring, employed to a higher degree by β-barrel proteins, seems to be due to both steric and electrochemical factors. In the interface region, the polar groups of these amino acids can interact with the phosphate groups, while the hydrophobic rings can interact with the lipid-chains. It has even been observed that there is a preference for Tyrosine and Tryptophan to point toward the phosphlipid headgroups, that is, residues that are located outside the interface region point the polar groups inward, while the ones located outside point the polar groups outward.[36] Phenylalanine is also aromatic, but entirely hydrophobic, and is not biased toward the interface region.

Charged or polar amino acids are energetically unfavorable within the membrane core. However, they are energetically tolerated toward the termini of the transmembrane helices.[26,36] In particular, the polar groups of the long side chains in arginine and lysine can "snorkel", that is, orient themselves so that the polar groups approach the interfacial and aqueous regions. This allows them to pull hydrating waters into the hydrocarbon part of the bilayer and create polar microenvironments for themselves. Obviously, a long side chain is advantageous for snorkeling and therefore Arginine is more frequent close to the membrane center than Lysine.[38]

Polar amino acids are also found in the center of the membrane when binding ions or lining water filled channels.[37] These residues are often functionally important and therefore evolutionarily conserved. Also, in contrast to polar amino acids found outside the membrane, within the membrane these residues are preferentially not exposed. This also partly contributes to their conservation, as it is more difficult to replace a buried amino acid than one that is exposed to the surrounding environment without causing the structure of the protein to change. However, even when taking this into account the polar residues are more conserved than other residues within the membrane core, indicating their functional importance.

Further, it can be noted that the positively charged (basic) amino acids Arginine and Lysine are more frequent on the inside (cytosolic) side of the membrane, confirming the so-called positive inside rule.[39,40] Positively charged, basic residues are strong topology signals with a preference for the cytosol. Acidic residues do not effect topology as strongly, or at all, and show no statistical preference for loops on either side of the membrane. However, they seem to have a greater influence on topology in Sec independent translocation. During Sec dependent, cotranslational insertion protein segments are affected by positively charged residues. If placed downstream of and close to a transmembrane segment of C_{in}-orientation, a single Arg or Lys can lower the apparent free energy of insertion of that segment with about 0.5 kcal/mol.[41] The effect is additive, the more charged residues, the lower the free energy. Also, a stretch of six lysines has been

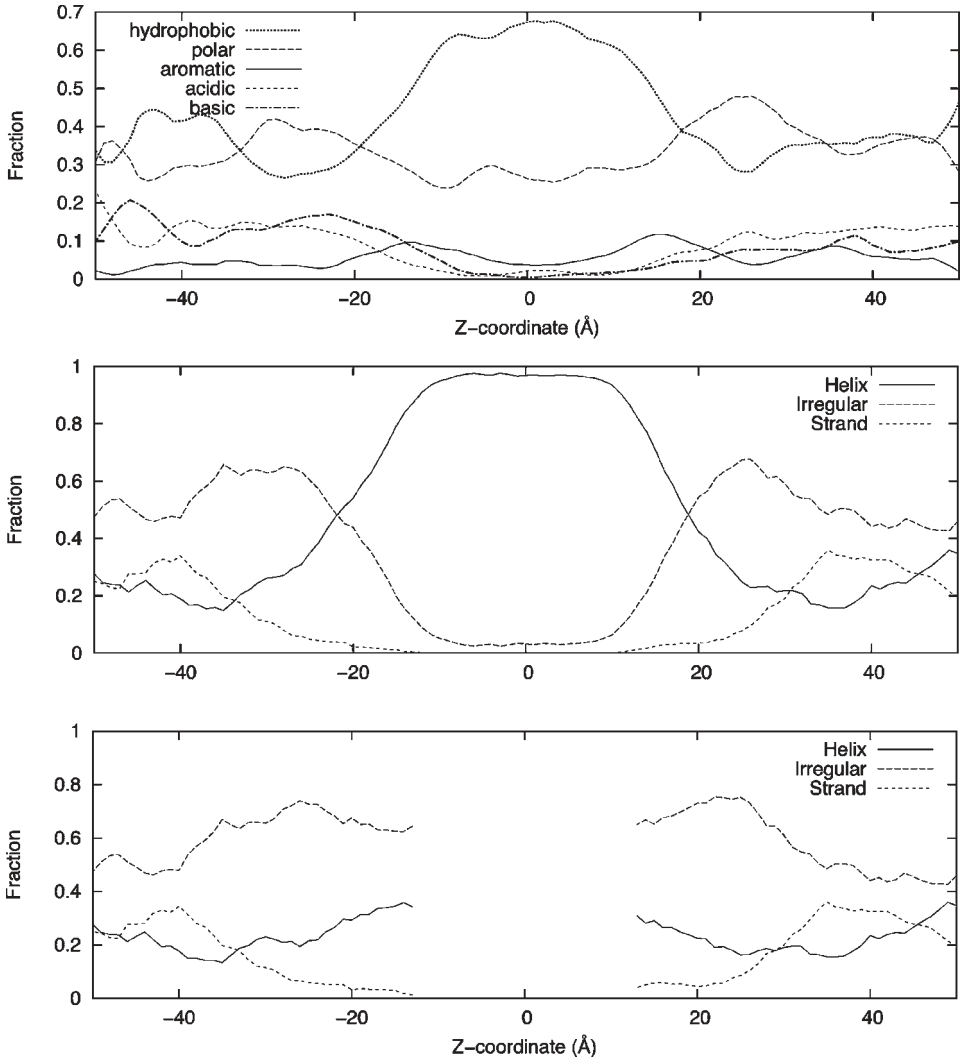

Figure 5. (a) Frequency of different amino acids along the membrane normal. (b) Secondary structure along the membrane normal. (c) Secondary structure frequencies when ignoring transmembrane helices.

shown to affect insertion from a distance of up to 25 residues away for a single-spanning protein and up to 12 residues away for a transmembrane helix in a multispanning protein, when expressed in an in vitro system.[42] Histidines are not charged at the physiological pH of the cell but have a similar effect if the pH is lowered.[43]

The exact mechanisms behind the positive inside rule are not yet fully understood, and they may differ slightly between Archaea, Bacteria and Eukarya. For example, the retaining effect of positive charges is more pronounced in *E. coli* than in microsomes (miniature eukaryotic ER membrane), most likely due to the much stronger electrochemical potential over the bacterial inner cell membrane.[44]

Secondary Structures within the Membrane

In general α-helical membrane proteins consist of a core of long, about 20 residues, helices, see Figure 5. In the vicinity of the phospholipid headgroups these helices are terminated and therefore this region is enriched in coil residues. It can also be noted that hardly any β-sheet residues are found close to the membrane. When taking a closer look it is clear that in addition to coil residues in this region it contains "interface-helices", but no β-sheets.[36]

Although most of the residues in the membrane core are in a helical state, about 7% coil residues are found.[33] This means that on average about one residue in each TM helix is in a nonhelical state. In a coil, the peptide backbone forms no regular secondary structure, hence exposing its polar backbone. It has been found that particularly in channels and transporters, this feature occurs frequently and is important for function. In TM helices, these coil residues might provide a higher degree of structural flexibility, creating swivels and hinges and also positioning side chains properly for interactions. As proposed by their conservation they are often functionally important. Both polar residues and backbone coil residues are most frequent among channels and transporters.

Reentrant Regions. As briefly mentioned above, reentrant loops are segments that penetrate the lipid bilayer without traversing it, entering and exiting on the same side. They have been estimated to occur in at least 10% of all multispanning membrane proteins, mostly in ion and water channels.[19,45] Although some attempts have been made to identify reentrant

dx.doi.org/10.1021/pr200145a |*J. Proteome Res.* 2011, 10, 3324–3331

regions from sequence alone no method can identify them with high confidence so these numbers are quite uncertain.[46] Most likely this is due to the large variation seen between different reentrant regions. Reentrant regions can either consist of (1) helix-turn-helix and are, if long enough, sometimes hard to distinguish from two transmembrane helices; (2) shorter segments of coil−helix or helix−coil, and (3) even shorter segments of purely irregular coil structure. Reentrant loops are enriched in prolines and small residues, especially glycines and alanines.

BIOINFORMATICS

The topology, that is, secondary structure, of membrane proteins is quite easy to predict even using simple methods. Some limited prediction accuracy of entire 3D-structures has also been reported.[47] The first methods only used hydrophobicity and were actually quite accurate. However, with the inclusion of the positive inside rule into the TOPPRED method the accuracy increased dramatically.[39] Slightly later, MEMSAT was introduced.[48] In contrast to earlier methods, the topology is predicted by finding the optimal path through a model mimicing a membrane proteins. The scoring is based on statistical preferences of amino-acids in different parts of transmembrane protein. This was much later used in methods based on hidden Markov models.[49,50] Surprisingly, recent benchmarks show that these two methods still perform quite well and often better than methods developed more recently. A clear improvement was not achieved until the appearance of methods combining multiple sequence alignments and hidden Markov models.[51] Recently, the inclusion of the more accurate biological hydrophobicity scale,[52] a combination of several machine learning methods,[46,53] and the use of consensus methods[54] have increased the accuracy even further.

Despite much progress, three major chanllenges for TM topology predictions remain. First, the appearance of more complex membrane protein structures has created a need for more complex methods to predict structural features of membrane proteins. One such method is OCTOPUS.[46] In OCTOPUS, a combination of artificial neural networks and hidden Markov models tries to identify not only the transmembrane regions of a membrane protein but also reentrant regions and interface helices. This has resulted in an increased accuracy in the topology predictions, but unfortunately the ability to identify nonstandard features is still quite limited. Most likely, this is due to the low number of membrane proteins with known structure. The second major challenge for membrane protein topology predictions is how to distinguish between membrane regions and signal peptides. The first method that could do this with some accuracy was Phobius[55] and later some improved methods have been introduced.[56−58] Finally, it has recently been shown that altough the prediction accuracy is similar between the best methods the predictions on a genome scale differs significantly.[59]

ASSOCIATED CONTENT

Supporting Information

Powerpoint presentation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: arne@bioinfo.se.

ACKNOWLEDGMENT

This work was supported by grants from the Swedish Research Council (VR-NT 2009-5072, VR-M 2010-3555). SSF (the Foundation for Strategic Research) through the CBR grant, The EU 6′Th Framework Program is gratefully acknowledged for support to the EMBRACE project, contract No: LSHG-CT-2004-512092. The EU 7′Th Framework Program is gratefully acknowledged for support to the EDICT project, contract No: FP7-HEALTH-F4-2007-201924. We thank Dr. Sara Light for proofreading.

REFERENCES

(1) Luckey, M. *Membrane Structural Biology - With Biochemical and Biophysical Foundations*; Cambridge University Press: New York, 2008.

(2) 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.

(3) Oberai, A.; Ihm, Y.; Kim, S.; Bowie, J. A limited universe of membrane protein families and folds. *Protein Sci.* **2006**, *15*, 1723–1734.

(4) Imai, K.; Gromiha, M.; Horton, P. Mitochondrial beta-barrel proteins, an exclusive club? *Cell* **2008**, *135*, 1158−1159; author reply 1159.60.

(5) Garrow, A.; Agnew, A.; Westhead, D. TMB-Hunt: an amino acid composition based method to screen proteomes for beta-barrel transmembrane proteins. *BMC Bioinform.* **2005**, *6*, 56.

(6) Elofsson, A.; von Heijne, G. Membrane protein structure: prediction versus reality. *Annu. Rev. Biochem.* **2007**, *76*, 125–140.

(7) 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.

(8) Wiener, M.; White, S. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys. J.* **1992**, *61*, 434–447.

(9) van Meer, G.; Voelker, D.; Feigenson, G. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 112–124.

(10) Koga, Y.; Nishihara, M.; Morii, H.; Akagawa-Matsushita, M. Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol. Rev.* **1993**, *57*, 164–182.

(11) von Heijne, G. Membrane-protein topology. *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 909–918.

(12) 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.

(13) Hedin, L.; Ojemalm, K.; Bernsel, A.; Hennerdal, A.; Illergard, K.; Enquist, K.; Kauko, A.; Cristobal, S.; von Heijne, G.; Lerch-Bader, M.; Nilsson, I.; Elofsson, A. Membrane Insertion of Marginally Hydrophobic Transmembrane Helices Depends on Sequence Context. *J. Mol. Biol.* **2009**, *1*, 221–229.

(14) Rehling, P.; Brandner, K.; Pfanner, N. Mitochondrial import and the twin-pore translocase. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 519–530.

(15) Emanuelsson, O.; Elofsson, A.; von Heijne, G.; Cristobal, S. In silico prediction of the peroxisomal proteome in fungi, plants and animals. *J. Mol. Biol.* **2003**, *330*, 443–456.

(16) Lee, P.; Tullman-Ercek, D.; Georgiou, G. The bacterial twin-arginine translocation pathway. *Annu. Rev. Microbiol.* **2006**, *60*, 373–395.

(17) Kauko, A.; Hedin, L.; Thebaud, E.; Cristobal, S.; Elofsson, A.; von Heijne, G. Repositioning of transmembrane alpha-helices during membrane protein folding. *J. Mol. Biol.* **2010**, *397*, 190–201.

(18) Pitonzo, D.; Skach, W. Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim. Biophys. Acta* **2006**, *1758*, 976–988.

(19) 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.
(20) Stevens, T.; Arkin, I. Turning an opinion inside-out: Rees and Eisenberg's commentary (Proteins 2000;38:121−122) on "Are membrane proteins 'inside-out' proteins?" (Proteins 1999;36:135−143). *Proteins* **2000**, *40*, 463–464.
(21) Adamian, L.; Liang, J. Prediction of transmembrane helix orientation in polytopic membrane proteins. *BMC Struct. Biol.* **2006**, *6*, 13.
(22) White, S.; Wimley, W. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
(23) Lins, L.; Brasseur, R. The hydrophobic effect in protein folding. *FASEB J.* **1995**, *9*, 535–540.
(24) Bowie, J. Membrane protein folding: how important are hydrogen bonds? *Curr. Opin. Struct. Biol.* **2010**, *11*, 42–49.
(25) Hessa, T.; Kim, H.; Bihlmaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S.; von Heijne, G. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **2005**, *433*, 377–381.
(26) 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.
(27) Johansson, A.; Lindahl, E. Protein contents in biological membranes can explain abnormal solvation of charged and polar residues. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15684–15689.
(28) Zhou, F.; Cocco, M.; Russ, W.; Brunger, A.; Engelman, D. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* **2000**, *7*, 154–160.
(29) Kim, S.; Jeon, T.; Oberai, A.; Yang, D.; Schmidt, J.; Bowie, J. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14278–14283.
(30) Adamian, L.; Liang, J. Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* **2002**, *47*, 209–218.
(31) Hildebrand, P.; Gunther, S.; Goede, A.; Forrest, L.; Frommel, C.; Preissner, R. Hydrogenbonding and packing features of membrane proteins: functional implications. *Biophys. J.* **2008**, *94*, 1945–1953.
(32) Gimpelev, M.; Forrest, L.; Murray, D.; Honig, B. Helical packing patterns in membrane and soluble proteins. *Biophys. J.* **2004**, *87*, 4075–4086.
(33) Kauko, A.; Illergard, K.; Elofsson, A. Coils in the membrane core are conserved and functionally important. *J. Mol. Biol.* **2008**, *380*, 170–180.
(34) Fu, D.; Libson, A.; Miercke, L.; Weitzman, C.; Nollert, P.; Krucinski, J.; Stroud, R. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **2000**, *290*, 481–486.
(35) 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. U.S.A.* **2004**, *101*, 959–963.
(36) 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.
(37) Illergard, K.; Kauko, A.; Elofsson, A. Why are polar residues within the membrane core evolutionary conserved? *Proteins* **2011**, *79*, 79–91.
(38) Johansson, A.; Lindahl, E. Position-resolved free energy of solvation for amino acids in lipid membranes from molecular dynamics simulations. *Proteins* **2008**, *70*, 1332–1344.
(39) von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **1992**, *225*, 487–494.
(40) von Heijne, G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* **1989**, *341*, 456–458.
(41) Lerch-Bader, M.; Lundin, C.; Kim, H.; Nilsson, I.; von Heijne, G. Contribution of positively charged flanking residues to the insertion of transmembrane helices into the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4127–4132.
(42) Kida, Y.; Morimoto, F.; Sakaguchi, M. Signal anchor sequence provides motive force for polypeptide chain translocation through the endoplasmic reticulum membrane. *J. Biol. Chem.* **2009**, *284*, 2861–2866.
(43) Andersson, H.; Bakker, E.; von Heijne, G. Different positively charged amino acids have similar effects on the topology of a polytopic transmembrane protein in Escherichia coli. *J. Biol. Chem.* **1992**, *267*, 1491–1495.
(44) Johansson, M.; Nilsson, I.; von Heijne, G. Positively charged amino acids placed next to a signal sequence block protein translocation more efficiently in Escherichia coli than in mammalian microsomes. *Mol. Gen. Genet.* **1993**, *239*, 251–256.
(45) Lasso, G.; Antoniw, J.; Mullins, J. A combinatorial pattern discovery approach for the prediction of membrane dipping (re-entrant) loops. *Bioinformatics* **2006**, *22*, e290–7.
(46) Viklund, H.; Elofsson, A. OCTOPUS: improving topology prediction by two-track ANNbased preference scores and an extended topological grammar. *Bioinformatics* **2008**, *24*, 1662–1668.
(47) Barth, P.; Wallner, B.; Baker, D. Prediction of membrane protein structures with complex topologies using limited constraints. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 1409–1414.
(48) Jones, D.; Taylor, W.; Thornton, J. A model recognition approach to the prediction of allhelical membrane protein structure and topology. *Biochemistry* **1994**, *33*, 3038–3049.
(49) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.
(50) Tusnady, G.; Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **2001**, *17*, 849–850.
(51) Vviklund, H.; Elofsson, A. Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci.* **2004**, *13*, 1908–1917.
(52) 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. U.S.A.* **2008**, *105*, 7177–7181.
(53) Jones, D. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **2007**, *23*, 538–544.
(54) Bernsel, A.; Viklund, H.; Hennerdal, A.; Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **2009**, *37*, W465–8.
(55) Kall, L.; Krogh, A.; Sonnhammer, E. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **2004**, *338*, 1027–1036.
(56) Viklund, H.; Bernsel, A.; Skwark, M.; Elofsson, A. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **2008**, *24*, 2928–2929.
(57) Nugent, T.; Jones, D. Transmembrane protein topology prediction using support vector machines. *BMC Bioinform.* **2009**, *10*, 159.
(58) Reynolds, S.; Kall, L.; Rife, M.; Bilmes, J.; Noble, W. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Comput. Biol.* **2008**, *4*, e1000213.
(59) Fagerberg, L.; Jonasson, K.; von Heijne, G.; Uhlen, M.; Berglund, L. Prediction of the human membrane proteome. *Proteomics* **2010**, *10*, 1141–1149.