Research Focus

The membrane–water interface region of membrane proteins: structural bias and the anti-snorkeling effect

Jie Liang, Larisa Adamian and Ronald Jackups Jr

Department of Bioengineering, University of Illinois at Chicago, M/C563, 835 S. Wolcott Avenue, Chicago, IL 60612-7340, USA

Membrane proteins have important roles in many cellular processes. Computational analysis of their sequences and structures has provided much insight into the organizing principles of transmembrane helices. In a recent study, the membrane–water interface region was examined in detail for the first time. The results have revealed that this interface region has an important role in constraining protein secondary structure. This study raises new questions and opens up new directions for studying membrane proteins.

Insights from computational analysis

Membrane proteins are abundant in most species and have important roles, including signal transduction, proton pumping, cell trafficking and photosynthesis. Understanding their structural organization and the principles governing their folding and assembly is an important task of biochemistry.

Computational analysis of membrane protein sequences has revealed fundamental insights. For instance, the success of the prediction of transmembrane (TM) helices from sequence hydropathy plots contributed to the formulation of the classic two-stage model of membrane-protein folding [1]. In addition, the observation of an asymmetric distribution of ionizable residues led to the discovery of the 'positive-inside' rule [2], by which arginine and lysine are four times more abundant in the cytoplasmic segments of membrane proteins than in the extracellular segments. Further insights were gained from analysis of the distribution [3] and sequence motifs [4] of amino acid types in the TM region.

Analysis of rapidly accumulating membrane-protein structures has been similarly fruitful. One example is the discovery of aromatic girdles, namely, the two belt regions of TM domains in which tryptophan and tyrosine are located in high proportions [5]. Studies of interacting helices have revealed the important roles of regular hydrogen-bond, weak hydrogen-bond and packing interactions in helical assembly [6–8], confirming earlier pioneering experimental studies [9,10]. Recent analysis of membrane-protein structures continues to reveal insights about TM helices, such as side-chain preferences and snorkeling effects [11,12].

Remarkably, results from computational analyses are largely in good agreement with experimental data. For example, the stabilities of amino acids in lipids inferred by computational analysis are consistent with experimental studies [13,14]. Recently, the sequence code for inserting a peptide into target membranes via the translocation machinery was deciphered [15]. In this study, the measured biological, physico-chemical and hydrophobicity scales all agree with each other, providing vital evidence that direct protein–lipid interactions are crucial for translocon-mediated membrane insertion.

The membrane–water interface region

Much has been learned about the TM region, but little is known about other regions of membrane proteins. The interface between the membrane and the aqueous solvent is a special boundary region that has different physico-chemical properties compared with either the lipid (TM helical) region or the bulk solvent. The crucial steps of membrane insertion occur here. What constraints does this region impose on the structure of membrane proteins? A recent study by Granseth *et al.* [16] is the first that brings this important region to the forefront of investigation.

Secondary structures in the interface region

The membrane–water interface region can be defined by the distance from the center of the membrane. Taking the region that is $\pm$15–25 Å from the center, Granseth *et al.* [16] analyzed 27 non-homologous protein structures containing 221 TM helices in total (Figure 1).

One of the main findings of Granseth *et al.* [16] is that the membrane–water interface region is dominated by irregular structures ($\sim$70%) and helices ($\sim$30%), but lacks β strands. The irregular structures are enriched with glycine and proline residues, which are well-known turn promoters and helix breakers.

In most cases, interface helices are connected to TM helices. Both types of helices are enriched with hydrophobic residues, but interface helices have a much higher content of polar aromatic residues (tryptophan and tyrosine). Frequently, a long peptide loop ($>$15 residues) connecting two TM helices contains an interface helix. There is little correlation between the end-to-end physical distance of two TM helices and the length of the connecting loop. This suggests that interface helices help to maintain the relative positions of TM helix ends while accommodating a large number of residues between helices.

Corresponding author: Liang, J. (jliang@uic.edu).
Available online 2 June 2005

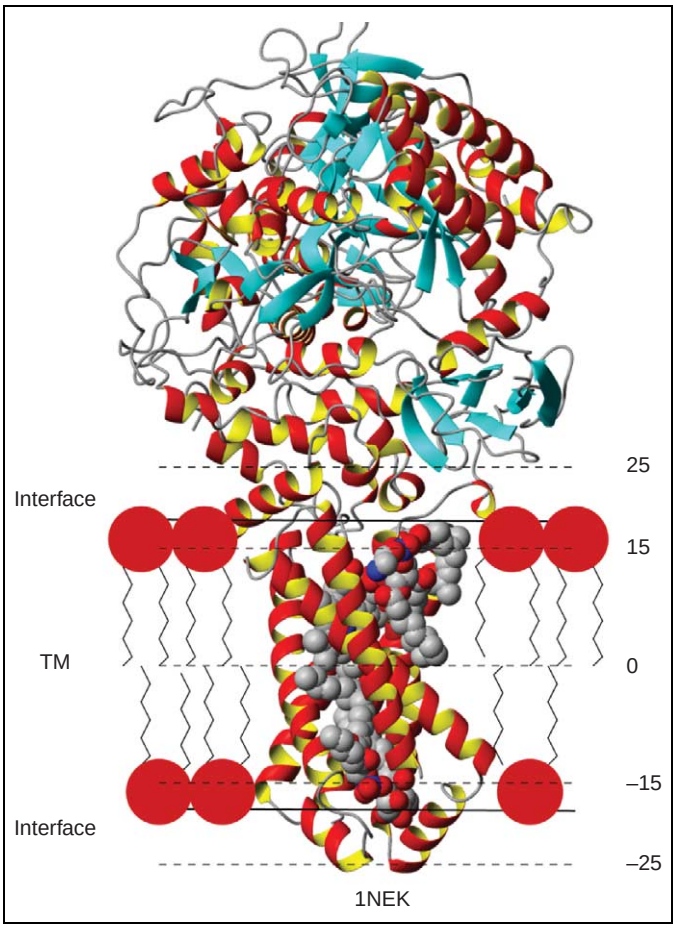

Figure 1. The spatial regions of helical membrane proteins. The transmembrane (TM) region is defined as the space from −15 Å to +15 Å in vertical distance perpendicular to the membrane center plane. The membrane–water interface is defined as the space between ±15 Å and ±25 Å. In the structure of succinate dehydrogenase (PDB code: 1NEK), the co-crystallized phospholipids cardiolipin and phosphatidylethanolamine (in full space filling) define the actual boundaries of the lipid bilayer (solid lines at ±18 Å). Structural elements that are contained in the ±15–25-Å region are interface helices and irregular structures. β strands are not found in the interface region.

Snorkeling and anti-snorkeling in membrane proteins

Residues such as lysine, arginine, tryptophan and tyrosine in TM helices often extend their side chains along the direction perpendicular to the membrane bilayer and point away from the membrane core. This is called the snorkeling effect [11]. For example, tyrosine can extend as much as 3.3 Å away from the C_β atom in this direction. By contrast, phenylalanine in the TM region tends to bend backwards, so that its aromatic ring is embedded in the hydrophobic core region. This is called the anti-snorkeling effect [11] (Figure 2a). However, in the interface region, Granseth *et al.* found that the side chains of tryptophan and tyrosine reverse orientation and tend to point towards the membrane core [16] (Figure 2b), thus changing from snorkeling to anti-snorkeling.

These observations can be explained by a general principle: the hydrophobic elements of a residue prefer to be embedded in the membrane hydrophobic core, whereas the polar elements tend to interact with the polar lipid head-group or aqueous environment [16,17]. The side chains of polar residues (e.g. lysine, arginine, asparagine and glutamine) tend to point away from the lipid hydrophobic core (snorkel) in both TM and interface regions. The side chains of hydrophobic residues (e.g. phenylalanine, leucine and isoleucine) tend to point towards the core region of the membrane (anti-snorkel) in both TM and interface regions. Amphipathic residues, such as tryptophan and tyrosine, which contain both hydrophobic and polar elements in their side chains, tend to snorkel away from the core when in the TM region, but reverse this direction and anti-snorkel when located in the interface region (Table 1).

Location bias

Side chains of amino acids can only take a few sterically allowed conformations [12]. This results in a variety of C- or N-terminal location biases: polar residues (e.g. lysine, arginine, asparagine and glutamine) are concentrated at

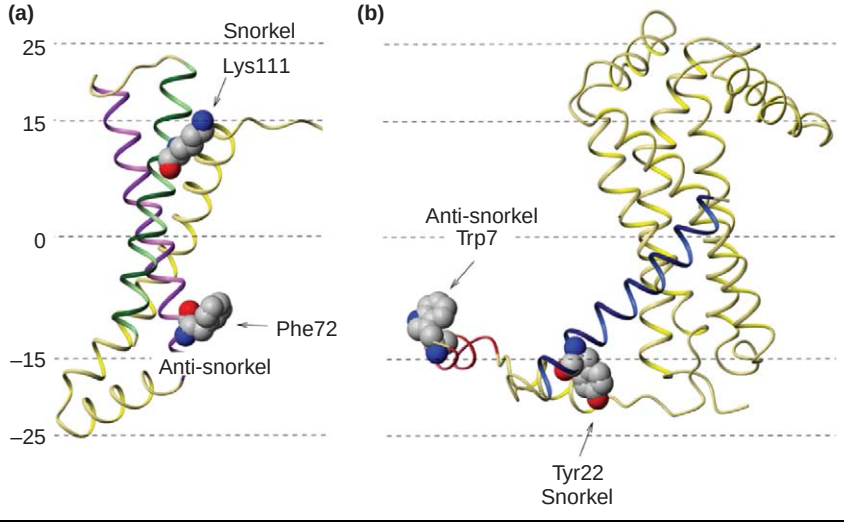

Figure 2. Examples of snorkeling and anti-snorkeling in membrane proteins. **(a)** The snorkeling behavior of polar and hydrophobic residues is illustrated by examples from subunit SdhC of succinate dehydrogenase (PDB code: 1NEK). Lys111 snorkels away from the core of the TM region, and Phe72 anti-snorkels towards the core of the TM region. **(b)** The snorkeling and anti-snorkeling behavior of amphipathic residues tyrosine and tryptophan is illustrated by examples from Cytochrome b_6 and PetL subunits of Cytochrome b_6f complex (PDB code: 1Q90). Tyr22 from the TM region of the PetL subunit snorkels away from the lipid hydrophobic core. The aromatic residue Trp7, which is located on an interface helix of Cytochrome *b*, anti-snorkels towards the core of the phospholipid bilayer.

Table 1. Summary of the snorkeling and anti-snorkeling behavior of different residues in the TM and interface regions

Region	Polar	Hydrophobic	Amphopathic
Transmembrane region	Snorkel	Anti-snorkel	Snorkel
Interface region	Snorkel	Anti-snorkel	Anti-snorkel

the N termini of helices, whereas hydrophobic residues (e.g. alanine, valine and isoleucine) and tyrosine are concentrated at the C termini. Because residue side chains tend to point towards the N termini in α helices, this makes N-terminal locations more favorable for polar residues to snorkel. Tyrosine has a special rotamer to enable its hydroxyl group to extend further from the core region, hence, tyrosine is more favored at the C termini. The biased locations of residues can be largely explained by the available side-chain rotamers and the propensity to snorkel.

Concluding remarks

The work of Granseth and colleagues opens up a new area for the study of membrane-protein biochemistry. Now, with the well-defined interface region and a clear picture of its constraints on structures of membrane proteins, we can start to ask new questions. For example, do interface helices form only in the constrained environment of the regions bordering the membrane? If the sequences of interface helices are introduced into soluble proteins, will they still form stable helical structures? Does the entropic effect of end-to-end distances for loops connecting two TM helices differ from that of soluble proteins? How can such thermodynamic considerations help to suggest mutants for enhanced stability or dynamics?

Undoubtedly, future studies of the interface region will facilitate understanding of the folding mechanisms of membrane proteins, and might lead to the development of engineering principles for designing novel and fully functional membrane proteins.

References

1 Popot, J.L. and Engelman, D.M. (2000) Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem.* 69, 881–922
2 von Heijne, G. (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* 5, 3021–3027
3 Arkin, I.T. and Brunger, A.T. (1998) Statistical analysis of predicted transmembrane α-helices. *Biochim. Biophys. Acta* 1429, 113–128
4 Senes, A. *et al*. (2000) Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with β-branched residues at neighboring positions. *J. Mol. Biol.* 296, 921–936
5 Weiss, M.S. *et al*. (1991) Molecular architecture and electrostatic properties of a bacterial porin. *Science* 254, 1627–1630
6 Adamian, L. and Liang, J. (2002) Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins* 47, 209–218
7 Eilers, M. *et al*. (2002) Comparison of helix interaction in membrane and soluble α-bundle proteins. *Biophys. J.* 82, 2720–2736
8 Senes, A. *et al*. (2001) The C_α–H…O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. U. S. A.* 98, 9056–9061
9 Choma, C. *et al*. (2000) Asparagine-mediated self-association of a model transmembrane helix. *Nat. Struct. Biol.* 7, 161–166
10 Zhou, F.X. *et al*. (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat. Struct. Biol.* 7, 154–160
11 Chamberlain, A.K. *et al*. (2004) Snorkeling preferences foster an amino acid composition bias in transmembrane helices. *J. Mol. Biol.* 339, 471–479
12 Chamberlain, A.K. and Bowie, J.U. (2004) Analysis of side-chain rotamers in transmembrane proteins. *Biophys. J.* 87, 3460–3469
13 Beuming, T. and Weinstein, H. (2004) A knowledge-based scale for the analysis and prediction of buried and exposed faces of transmembrane domain proteins. *Bioinformatics* 20, 1822–1835
14 Adamian, L. *et al*. (2005) Empirical lipid propensities of amino acid residues in multispan α helical membrane proteins. *Proteins* 59, 496–509
15 Hessa, T. *et al*. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433, 377–381
16 Granseth, E. *et al*. (2005) A study of the membrane-water interface region of membrane proteins. *J. Mol. Biol.* 346, 377–385
17 White, S.H. and Wimley, W.C. (1999) Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28, 319–365

0968-0004/$ - see front matter
doi:10.1016/j.tibs.2005.05.003

Testis-specific histone H3 expression in somatic cells

Jérôme Govin, Cécile Caron, Sophie Rousseaux and Saadi Khochbin

INSERM U309 – Laboratoire de Biologie Moléculaire et Cellulaire de la Différenciation Équipe Chromatine et expression des gènes Institut Albert Bonniot Faculté de Médecine, Domaine de la Merci 38706 La Tronche Cedex, France

Histone variants functionally differentiate individual nucleosomes and, hence, act as key regulators of chromatin structure and function. Large-scale proteomic projects are now valuable sources of histone-variant discovery, showing, in particular, that somatic mammalian cells express a larger panel of histone H3 variants than previously thought, including testis-specific variants and as yet uncharacterized species. These data also suggest a tight relationship between the complexity of histone-variant expression and physiopathological states of the cells.

Corresponding author: Khochbin, S. (khochbin@ujf-grenoble.fr).
Available online 26 May 2005