BMB
reports

Mini Review

Bioinformatic approaches for the structure and function of membrane proteins

Hyun-Jun Nam[1], *Jouhyun Jeon*[2] *& Sanguk Kim*[1,2,*]

[1]School of Interdisciplinary Bioscience and Bioengineering, [2]Division of Molecular and Life Science, Pohang University of Science and Technology, Pohang 790-784, Korea

Membrane proteins play important roles in the biology of the cell, including intercellular communication and molecular transport. Their well-established importance notwithstanding, the high-resolution structures of membrane proteins remain elusive due to difficulties in protein expression, purification and crystallization. Thus, accurate prediction of membrane protein topology can increase the understanding of membrane protein function. Here, we provide a brief review of the diverse computational methods for predicting membrane protein structure and function, including recent progress and essential bioinformatics tools. Our hope is that this review will be instructive to users studying membrane protein biology in their choice of appropriate bioinformatics methods. [BMB reports 2009; 42(11): 697-704]

INTRODUCTION

Membrane proteins play important roles in many physiological processes, constituting approximately 20-30% of all open reading frames (ORFs) in fully sequenced genomes. Membrane proteins are involved in various functions, such as intercellular communication, molecular transport and biogenesis, and are associated with various human diseases, such as Alzheimer's disease, diabetes, Hodgkin's disease and liver cirrhosis. Many membrane proteins are primary drug targets because their localization is easy for drug delivery and their functions are mainly involved in converting extracellular signals into intracellular processes. Currently, about 60% of approved drugs target membrane proteins (1). For example, G protein-coupled receptors (GPCRs) account for 25% of all drug targets on the market, and nuclear hormone receptors and ion channels account for 7%. Therefore, elucidating the structure and function of membrane proteins is essential for the identification of drug targets and development of therapeutic applications.

To better understand protein function, biologists have focused on the determination of high-resolution protein structure, which is necessary for the development of drugs that specifically block ligand binding sites or inhibit protein oligomerization (2). Even so, the determination of high-resolution protein structure has been a bottleneck to the characterization of membrane protein function for many years due to difficulties in protein expression, purification and crystallization. Consequently, membrane protein structures represent less than 1% of the Protein Data Bank (PDB) (3), and only 70 high-resolution membrane protein structures exist compared to the thousands of water-soluble protein structures (4). Moreover, high-throughput function screenings, such as yeast two-hybrid and liquid chromatography-tandem mass spectrometry, cannot be easily applied to due to the hydrophobic nature of membrane proteins (5-7). Therefore, computational methods for predicting membrane protein structure and function could provide alternative solutions.

In this review, we focus on computational methods for predicting membrane protein structure and function, discussing recent progress as well as future developments. Computational methods for predicting membrane protein topology are classified as either knowledge-based or physical potential-based. Recent progress has been made in the development of machine learning techniques and in the integration of evolutionary information. Other topics discussed include advancements in bioinformatic approaches for the identification of residues that are functionally important to membrane proteins. In Fig. 1, we present an overview that describes how the methods discussed in this review could be applied to study the membrane protein structure and function.

Early developments in the prediction of membrane protein topology

Prediction of membrane protein topology, especially of transmembrane (TM) segments, is an important first step toward the understanding of membrane protein structure. Generally, membrane proteins span the membrane bilayer via unbroken helices that lead to limited topology. Therefore, if the locations of TM regions and loops are known, it is possible to predict the topology of membrane proteins from sequences only.

Membrane protein topology prediction algorithms are based on two basic observations of α-helical membrane proteins. TM segments are mainly composed of hydrophobic residues as they

*Corresponding author. Tel: 82-54-279-2348; Fax: 82-54-279-2199; E-mail: sukim@postech.ac.kr

Received 29 September 2009

Keywords: Evolutionary approaches, Function prediction, Membrane protein, Protein structure, Topology prediction

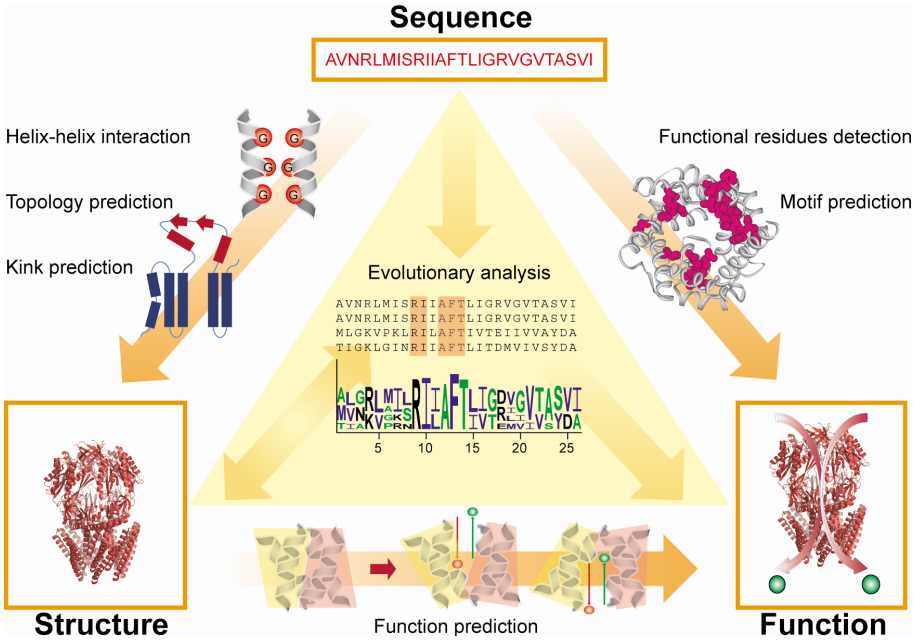

Fig. 1. Evolutionary analysis is used for predicting both the structure and function of membrane proteins. Predictions of topology and helix-helix interactions based on the detection of GxxxG motifs and kinks in amino acid sequences are important in the study of membrane protein structure. Detection of functional residues and prediction of motifs can also provide insights into the understanding of membrane protein function.

span the hydrophobic interior of a biological membrane, which is composed of two layers of lipid molecules, i.e. phospholipid bilayer. Phospholipids contain polar heads that interact with water molecules on both sides of membrane, and their hydrophobic fatty acid regions form a thick hydrophobic core. Membrane proteins have adapted to this environment by evolving mostly non-polar and hydrophobic amino acids in their TM segments (8, 9). Kyte and Doolittle developed the first method to predict TM segments based on this observation (10). They scanned a protein sequence with a sliding window of 19 amino acids and identified TM segments 20 to 30 residues in length based on overall hydrophobicity (11, 12). For more accurate predictions, others attempted to optimize hydrophobic scales, window sizes, and amino-acid propensities (13-15). For example, the aromatic residues Trp and Tyr tend to be located near the ends of TM segments, an observation which allowed the prediction of membrane protein topology to be improved (16, 17). Researchers also described the 'positive-inside rule', in that more positively charged residues, such as Lys and Arg, are found in the short cytoplasmic loops that connect TM segments compared to extracellular loops (18). These observations allowed the prediction of TM segment orientation for membrane proteins. Furthermore, integration of TM segment sequence features as well as increased numbers of available membrane protein structures have further improved TM topology prediction (19-23).

Recent progress in the prediction of membrane protein topology

Increased numbers of high-resolution membrane protein structures and the advancement of computational methods have increased the accuracy of membrane protein topology prediction. Most predictions are based on machine-learning techniques, such as neural networks (NNs) (24) and hidden Markov models (HMMs) (25, 26). Prediction methods using NNs, the first of which was PHDhtm (27), can give the probability of each residue being located in a TM segment, and therefore can define local sequence patterns as TM regions. Meanwhile, methods that implement HMMs complement NNs by providing more specialized models. HMM methods can capture the global sequence patterns of a membrane protein sequence, including repeats of TM segment-loop-TM segment patterns in α-helical membrane proteins. HMM was implemented in TMHMM (28) and HMMTOP (25) for TM topology prediction.

Methods for the prediction of TM segments have their own unique advantages. Several studies have compared the sensitivities of various methods to predict membrane protein topology (29-33). Jonathan and colleagues examined the performance of 9 different TM segment prediction methods using various sensitivity measures, such as number of TM segments, position of TM segments, and orientation of TM segments (32). They compared performance and found that no method was consistently the best. SPLIT4 (22), TMHMM2 (26), HMMTOP2 (34) and TMAP (35) predicted the number and position of TM segments well with accuracies over 80%. Meanwhile, TMHMM2, MEMSAT2 (36) and SPLIT4 were the best methods for predicting the start and end points of TM segments. TMHMM and SOSUI (21) were the most accurate methods for discriminating membrane proteins from soluble proteins with accuracies over 97% (29). Taken together, these methods have

different sensitivities for predicting the number and positions of TM helices, therefore helping researchers choose appropriate tools depending on their purpose.

Recently, several approaches have been developed to improve the prediction of membrane protein topology through a consensus prediction method (37-41). Consensus prediction methods are based on a decision making process that seeks agreement among most participants and resolves or mitigates minority objections. Consensus methods furthermore integrate and optimizate data from other prediction methods in order to obtain more reliable TM topologies (37). ConPred II combined the results of eight prediction methods, KKD (20), TMpred (23), TopPred II (42), DAS (43), TMAP (35), MEMSAT 1.8 (36), SOSUI (21), TMHMM 2 (26) and HMMTOP 2 (34), into a joint prediction histogram in order to more reliablely predict the location of TM segments (40). As a result, ConPred II improved prediction of TM topology and position compared to individual prediction methods. Another consensus method, TOPCONS (41), integrates the results of 5 different prediction methods: OCTOPUS (44), PRO-TMHMM, PRODIV-TMHMM (45), SCAMPI-single and SCAMPI-multi (46). TOPCONS outperformed all other methods in a benchmark test evaluating individual and consensus prediction methods. The improved predictions provided by the consensus methods are due to integration of multiple predictors. Thus, improvement in the accuracy of an individual predictor will likewise improve the accuracy of the overall consensus prediction method. In Table 1, we compiled the available TM prediction methods and their web addresses, and compared their functionality for user convenience.

Evolutionary approaches and data integration

Evolutionary information has been applied to the prediction of membrane protein topology. Several benchmark studies found that evolutionary approaches, such as poly-Phobius (47), MEMSAT3 (48), HMMTOP, prodiv-TMHMM and MemBrain (49), significantly improved the accuracy of TM segment prediction compared to conventional methods. Evolutionary approaches perform multiple sequence alignment in order to calculate the best average path to find correct topology. Other types of evolutionary approaches, such as phylogenetic analysis (50-52), characterize the functional mechanism of membrane proteins. Phylogenetic methods operate on the assumption that proteins associated together in a pathway or structural complex evolve in a correlated fashion, which helps determine the functional annotation of uncharacterized membrane proteins.

Data-based modeling combines biochemical and computational methods with low-resolution structural data in order to build a low-resolution model of membrane proteins. Various biochemical and spectroscopic experiments, such as solid-state NMR (ssNMR) (53-57), electron microscopy (EM) (58-60) and infrared spectroscopy (IR spectroscopy) (61-63), have rapidly accumulated low-resolution, two-dimensional structural information on membrane proteins. Computational methods have been developed to improve membrane protein structural models when high-resolution structural information is not available. Moreover, data-based modeling provides additional information on the conformational changes and dynamics of membrane proteins that are not easily obtained from x-ray crystal structures.

Identification of functionally important residues from the membrane protein sequence

Large-scale genome sequencing projects require the development of computational methods that can efficiently identify biological functions of proteins at the genome-wide level. Structural information on membrane proteins can provide important clues for elucidating functional mechanisms, but structure is difficult to obtain due to experimental issues associated

Table 1. Transmembrane prediction methods

Methods	URL	Additional options				
		Signal-peptide prediction	N-terminal orientation	Large-scale analysis	Freely downloadable	Kingdom specificity
DAS-Tmfilter	http://www.enzim.hu/DAS/DAS.html	x	x	o	x	x
HMMTOP2	http://www.enzim.hu/hmmtop	x	o	x	o	x
MEMSAT3	http://bioinf.cs.ucl.ac.uk/psipred/psiform.html	o	o	x	o	x
MINNOU	http://minnou.cchmc.org	x	x	x	o	x
PHDhtm	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_htm.html	x	x	o	x	x
Phobius	http://phobius.sbc.su.se	o	o	x	o	x
SOSUI	http://bp.nuap.nagoya-u.ac.jp/sosui	x	x	o	x	x
SPLIT4	http://split.pmfst.hr/split/4	x	x	x	x	x
TMHMM2	http://www.cbs.dtu.dk/services/TMHMM	x	o	o	o	x
Conpred2	http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2	o	o	o	x	o
TOPCONS	http://topcons.net	x	o	x	x	x
MemBrain	http://www.csbio.sjtu.edu.cn/bioinf/MemBrain	o	x	x	x	o

with sample preparation. Therefore, various sequence-based methods have been developed for the prediction of membrane protein function in addition to topology.

Helix-packing motifs in TM helices

TM helix-packing motifs play important structural and functional roles in membrane proteins (64). Therefore, their identification is essential for modeling of membrane protein structures. The most well characterized sequence signature that mediates TM helix-packing is the GxxxG motif (65, 66). The GxxxG motif, composed of 2 glycines at positions *i* and *i* + 4 on the same side of the helix, is statistically overrepresented in TM helices of membrane proteins. Furthermore, the glycine zippers, (G,A,S)xxxGxxxG and GxxxGxxx(G,S,T) contain multiple GxxxG motifs and are shown to have important functional roles in many membrane proteins (66). Due to their small size, the glycines in these motifs can facilitate the formation of helix-helix structures by optimizing van der Waals interactions or establishing hydrogen bonds between residues. Moreover, mutations in these motifs are shown to have deleterious effects on the function of membrane proteins (66-68).

Z-coordinate predictions of TM helices in membrane proteins

Z-coordinates, the distances between residues and the center of the membrane, of alpha-helical membrane proteins contain structural information on TM helix kinks, reentrant helices, interfacial helices, and loop protrusion from the membrane. TM helix kinks commonly occur at proline residues (69) where they mediate helix distortions and facilitate helix movements, which are essential for the operation of TM channels (70-72). TM helix kinks also assist protein folding by preventing off-pathway intermediates (73). Yohannan and colleagues found through helix kink predictions that the structural diversity of GPCRs is due to unique helix kink patterns, which once elucidated advance the understanding of GPCR function (64). Meanwhile, interfacial helices and reentrant loops, in which the polypeptide dips only partway across the membrane, contain important information on membrane protein function, such as the selectivity filter and gates of membrane pores (74).

Common occurrence of internal repeat symmetry in membrane proteins

The sequence of a protein determines both its structure and overall function. It has been suggested that symmetry is important for membrane proteins that undergo two-state allosteric transformations, such as active/inactive transitions or large conformational changes in transporters and channels (75). Internal repeat symmetry in a protein sequence is manifested by the association of similar subunits or by gene duplication events (76). Recently, we reported the common occurrence of internal repeat symmetry in membrane proteins and devised a method to identify internal repeats in membrane protein sequences (Fig. 2) (77). Remarkably, a genome-wide search of internal repeats in the sequences of membrane proteins revealed that nearly half of all membrane proteins contain symmetric internal repeats. Interestingly, most symmetric membrane proteins were found to be transporters, suggesting that symmetry may be beneficial for the large conformational changes during the substrate transport mechanism. Moreover, the most conserved motifs in internal repeats were found at the symmetric interfaces of protein structures in which functionally important residues of membrane proteins are frequently located.

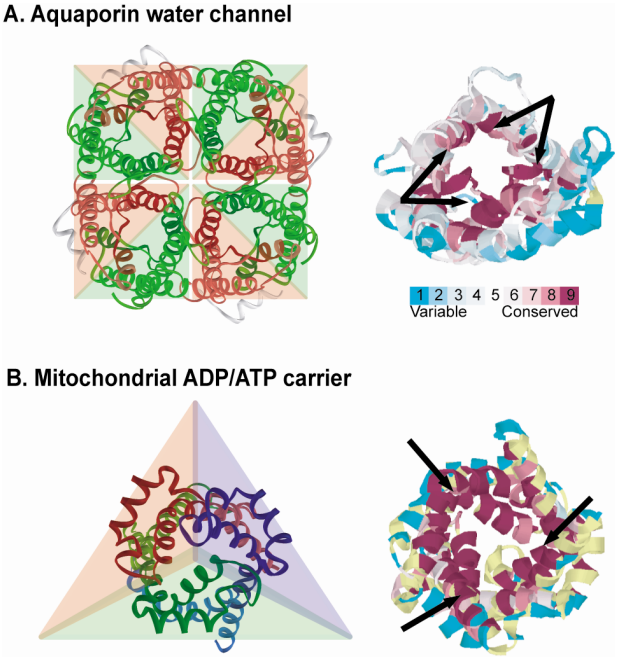

Fig. 2. Topological patterns and evolutionary rates of symmetric membrane proteins. (A) Aquaporin water channel (PDB ID: 1J4N) and (B) Mitochondrial ADP/ATP carrier (PDB ID: 1OKC) have 2-fold inverted symmetry and 3-fold symmetry, respectively. Each triangle represents a symmetric unit (left). Evolutionary rates were colored onto the symmetric membrane structures (right). The color-coding bar shows the site-specific evolutionary rate. Red indicates the conserved site and blue indicates the variable site. Black arrow represents the most conserved regions in the structure.

Combining evolutionary information for the identification of functionally important residues in membrane proteins

Combining evolutionary features from sequence information has been used to successfully identify functionally important residues in membrane proteins (78). Membrane protein transporters are involved in two-state allosteric communication, which mediates the propagation of regulatory information

from the substrate-binding site to the translocation pathway through large conformational changes mediated by cooperative residues. Recent studies have suggested that cooperative residues tend to be conserved (79) or evolutionarily coupled (80) in order to maintain allosteric communication. We recently reported a new method for the identification of cooperative residues of membrane protein transporters by integrating two different evolutionary features, sequence conservation and co-evolutionary information (78). The results demonstrated that conserved cores of evolutionarily coupled residues are involved in substrate recognition and translocation of membrane protein transporters. Moreover, a subset of these newly identified residues forms an interaction network connecting functional sites in the protein structure. The identification of functionally important residues from sequences provides an alternative method in which the function of membrane proteins can be discovered.

Current practice and future development

Prediction methods for membrane protein topology have improved understanding of membrane protein structure and function over recent decades. Despite recent progress in TM protein topology prediction, several limitations still exist. Firstly, since signal peptide sequences contain hydrophobic amino acids also found in TM segments, topology prediction methods will often falsely predict a signal peptide as a putative TM segment (81). This makes it difficult to discriminate between membrane proteins and soluble proteins with a signal peptide, significantly hampering the prediction of single-pass membrane proteins. To overcome this, advanced methods were developed by assigning additional N-terminal signal peptides using signal peptide prediction tools and by selecting non-overlapping TM segments separate from the signal peptide (82). Phobius and MEMSAT3 adopted this method and significantly improved accuracy of TM topology prediction (48, 83).

Secondly, many topology prediction methods have a problem in defining the exact length of TM helices. Topology predictors typically search for TM helices 17-25 amino acids in length. However, the lengths of TM helices in x-ray crystal structures range from 10 to 40 amino acids (29, 32). Furthermore, TM segments can be tilted at various angles or broken inside the lipid bilayer, resulting in reentrant loops and half-TM helices (84, 85). TM segments of various lengths can also cause errors in when predicting the number of TM segments. For example, a TM segment 36 amino acids in length would instead be predicted as two TM segments composed of 18 amino acids each. Such an error may affect the overall prediction of TM protein topology. To overcome this, the distance from the center of the membrane to each residue was measured. Granseth and colleagues have developed an algorithm to correctly identify reentrant loops and half-TM helices using this method (86).

Finally, the prediction of membrane protein function is another important issue. Many sequence-based or structure-based prediction methods were developed to determine the function of membrane proteins. However, prediction of membrane protein function from homologues sequences remains difficult due to the various functions of homologous proteins. Moreover, the three-dimensional structure of a membrane protein does not reveal much about its function unless accompanied by experimental results. Therefore, the development of comparative genomic approaches provides an alternative approach for the prediction of membrane protein function. These approaches analyze the conservation patterns of membrane protein family members in order to characterize the function of unannotated protein sequences (77). Information on protein-protein interactions and phylogenetic analysis were also used in the determination of protein function (50, 87).

Conclusions

Membrane proteins play important roles in the biology of the cell, including intercellular communication and molecular transport. Accurate prediction of membrane protein topology can provide a framework for the understanding of membrane protein structure and function. For example, various computational methods for predicting membrane protein topology can be immediately applied. Prediction methods have become more accurate and useful over the years although issues remain concerning single-pass TM proteins and the positions of TM helices. Nevertheless, computational methods can guide experiment planning and dataset interpretation by providing information on TM helix topology, helix-helix interactions, and functionally important regions of membrane proteins. Moreover, the integration of evolutionary information can increase the understanding of the structure and function of membrane proteins (77). We expect that rapid progress in the development of bioinformatic approaches will expand our knowledge of the sequence-structure-function relationship of membrane proteins.

Acknowledgements

This work was supported by the Korea Science and Engineering Foundation grant (M10753020006-07N5302-00610 and R01-2007-000-20425-0) and the World Class University program (R31-2008-000-10100-0).

REFERENCES

1. Hopkins, A. L. and Groom, C. R. (2002) The druggable genome. *Nat. Rev. Drug. Discov.* **1**, 727-730.
2. Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., de Vos, A. M. and Sliwkowski, M. X. (2004) Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5**, 317-328.
3. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The protein data bank. *Nucleic. Acids Res.* **28**, 235-242.
4. White, S. H. (2004) The progress of membrane protein structure determination. *Protein Sci.* **13**, 1948-1949.

5. Thaminy, S., Auerbach, D., Arnoldo, A. and Stagljar, I. (2003) Identification of novel ErbB3-interacting factors using the split-ubiquitin membrane yeast two-hybrid system. *Genome Res.* **13**, 1744-1753.
6. Santoni, V., Molloy, M. and Rabilloud, T. (2000) Membrane proteins and proteomics: un amour impossible? *Electrophoresis* **21**, 1054-1070.
7. Buttner, K., Bernhardt, J., Scharf, C., Schmid, R., Mader, U., Eymann, C., Antelmann, H., Volker, A., Volker, U. and Hecker, M. (2001) A comprehensive two-dimensional map of cytosolic proteins of Bacillus subtilis. *Electrophoresis* **22**, 2908-2935.
8. Bowie, J. U. (2005) Solving the membrane protein folding problem. *Nature* **438**, 581-589.
9. White, S. H., Ladokhin, A. S., Jayasinghe, S. and Hristova, K. (2001) How membranes shape protein structure. *J. Biol. Chem.* **276**, 32395-32398.
10. Kyte, J. and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
11. von Heijne, G. and Blomberg, C. (1979) Trans-membrane translocation of proteins. The direct transfer model. *Eur. J. Biochem.* **97**, 175-181.
12. Engelman, D. M. and Steitz, T. A. (1981) The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis. *Cell* **23**, 411-422.
13. Engelman, D. M., Steitz, T. A. and Goldman, A. (1986) Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 321-353.
14. Wimley, W. C. and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **3**, 842-848.
15. Jayasinghe, S., Hristova, K. and White, S. H. (2001) Energetics, stability, and prediction of transmembrane helices. *J. Mol. Biol.* **312**, 927-934.
16. Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G. and Elofsson, A. (1997) Architecture of helix bundle membrane proteins: an analysis of cytochrome c oxidase from bovine mitochondria. *Protein Sci.* **6**, 808-815.
17. Weiss, M. S., Kreusch, A., Schiltz, E., Nestel, U., Welte, W., Weckesser, J. and Schulz, G. E. (1991) The structure of porin from Rhodobacter capsulatus at 1.8 A resolution. *FEBS Lett.* **280**, 379-382.
18. von Heijne, G. (1992) Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487-494.
19. Zhou, H. and Zhou, Y. (2003) Predicting the topology of transmembrane helical proteins using mean burial propensity and a hidden-Markov-model-based method. *Protein Sci.* **12**, 1547-1555.
20. Klein, P., Kanehisa, M. and DeLisi, C. (1985) The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta.* **815**, 468-476.
21. Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14**, 378-379.
22. Juretic, D., Zoranic, L. and Zucic, D. (2002) Basic charge clusters and predictions of membrane protein topology. *J. Chem. Inf. Comput. Sci.* **42**, 620-632.
23. Hofmann, K. and Stoffel, W. (1993) TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.
24. Rost, B., Casadio, R., Fariselli, P. and Sander, C. (1995) Transmembrane helices predicted at 95% accuracy. *Protein Sci.* **4**, 521-533.
25. Tusnady, G. E. and Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Biol.* **283**, 489-506.
26. Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567-580.
27. Rost, B., Fariselli, P. and Casadio, R. (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* **5**, 1704-1718.
28. Sonnhammer, E. L., von Heijne, G. and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175-182.
29. Chen, C. P., Kernytsky, A. and Rost, B. (2002) Transmembrane helix predictions revisited. *Protein Sci.* **11**, 2774-2791.
30. Ikeda, M., Arai, M., Lao, D. M. and Shimizu, T. (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In Silico. Biol.* **2**, 19-33.
31. Jayasinghe, S., Hristova, K. and White, S. H. (2001) MPtopo: a database of membrane protein topology. *Protein Sci.* **10**, 455-458.
32. Cuthbertson, J. M., Doyle, D. A. and Sansom, M. S. (2005) Transmembrane helix prediction: a comparative evaluation and analysis. *Protein Eng. Des. Sel.* **18**, 295-308.
33. Moller, S., Croning, M. D. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646-653.
34. Tusnady, G. E. and Simon, I. (2001) The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849-850.
35. Persson, B. and Argos, P. (1997) Prediction of membrane protein topology utilizing multiple sequence alignments. *J. Protein Chem.* **16**, 453-457.
36. Jones, D. T., Taylor, W. R. and Thornton, J. M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* **33**, 3038-3049.
37. Nilsson, J., Persson, B. and von Heijne, G. (2000) Consensus predictions of membrane protein topology. *FEBS Lett.* **486**, 267-269.
38. Nilsson, J., Persson, B. and Von Heijne, G. (2002) Prediction of partial membrane protein topologies using a consensus approach. *Protein Sci.* **11**, 2974-2980.
39. Xia, J. X., Ikeda, M. and Shimizu, T. (2004) ConPred_elite: a highly reliable approach to transmembrane topology predication. *Comput. Biol. Chem.* **28**, 51-60.
40. Arai, M., Mitsuke, H., Ikeda, M., Xia, J. X., Kikuchi, T., Satake, M. and Shimizu, T. (2004) ConPred II: a consensus prediction method for obtaining transmembrane

topology models with high reliability. *Nucleic. Acids. Res.* **32**, W390-393.
41. Bernsel, A., Viklund, H., Hennerdal, A. and Elofsson, A. (2009) TOPCONS: consensus prediction of membrane protein topology. *Nucleic. Acids. Res.* **37**, w465-468.
42. Claros, M. G. and von Heijne, G. (1994) TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. BioSci.* **10**, 685-686.
43. Cserzo, M., Wallin, E., Simon, I., von Heijne, G. and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng.* **10**, 673-676.
44. Viklund, H. and Elofsson, A. (2008) OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**, 1662-1668.
45. Viklund, H. and Elofsson, A. (2004) Best alpha-helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci.* **13**, 1908-1917.
46. Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G. and Elofsson, A. (2008) Prediction of membrane-protein topology from first principles. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 7177-7181.
47. Kall, L., Krogh, A. and Sonnhammer, E. L. (2005) An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics 21 Suppl* **1**, i251-257.
48. Jones, D. T. (2007) Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538-544.
49. Shen, H. and Chou, J. J. (2008) MemBrain: improving the accuracy of predicting transmembrane helices. *PLoS ONE* **3**, e2399.
50. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. and Yeates, T. O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4285-4288.
51. Shimizu, T., Mitsuke, H., Noto, K. and Arai, M. (2004) Internal gene duplication in the evolution of prokaryotic transmembrane proteins. *J. Mol. Biol.* **339**, 1-15.
52. Liu, Y., Gerstein, M. and Engelman, D. M. (2004) Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 3495-3497.
53. Nishimura, K., Kim, S., Zhang, L. and Cross, T. A. (2002) The closed state of a H+ channel helical bundle combining precise orientational and distance restraints from solid state NMR. *Biochemistry* **41**, 13170-13177.
54. Kim, S., Quine, J. R. and Cross, T. A. (2001) Complete cross-validation and R-factor calculation of a solid-state NMR derived structure. *J. Am. Chem. Soc.* **123**, 7292-7298.
55. Wang, J., Kim, S., Kovacs, F. and Cross, T. A. (2001) Structure of the transmembrane region of the M2 protein H(+) channel. *Protein Sci.* **10**, 2241-2250.
56. Kim, S. and Cross, T. A. (2002) Uniformity, ideality, and hydrogen bonds in transmembrane alpha-helices. *Biophys. J.* **83**, 2084-2095.
57. Opella, S. J. and Marassi, F. M. (2004) Structure determination of membrane proteins by NMR spectroscopy. *Chem. Rev.* **104**, 3587-3606.
58. Fleishman, S. J., Unger, V. M., Yeager, M. and Ben-Tal, N. (2004) A Calpha model for the transmembrane alpha helices of gap junction intercellular channels. *Mol. Cell.* **15**, 879-888.
59. Fleishman, S. J., Harrington, S. E., Enosh, A., Halperin, D., Tate, C. G. and Ben-Tal, N. (2006) Quasi-symmetry in the cryo-EM structure of EmrE provides the key to modeling its transmembrane domain. *J. Mol. Biol.* **364**, 54-67.
60. Fleishman, S. J., Unger, V. M. and Ben-Tal, N. (2006) Transmembrane protein structures without X-rays. *Trends Biochem. Sci.* **31**, 106-113.
61. Arbely, E., Kass, I. and Arkin, I. T. (2003) Site-specific dichroism analysis utilizing transmission FTIR. *Biophys. J.* **85**, 2476-2483.
62. Mukherjee, P., Kass, I., Arkin, I. T. and Zanni, M. T. (2006) Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3528-3533.
63. Mukherjee, P., Kass, I., Arkin, I. T. and Zanni, M. T. (2006) Structural disorder of the CD3zeta transmembrane domain studied with 2D IR spectroscopy and molecular dynamics simulations. *J. Phys. Chem. B.* **110**, 24740-24749.
64. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. and Bowie, J. U. (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 959-963.
65. MacKenzie, K. R., Prestegard, J. H. and Engelman, D. M. (1997) A transmembrane helix dimer: structure and implications. *Science* **276**, 131-133.
66. Kim, S., Jeon, T. J., Oberai, A., Yang, D., Schmidt, J. J. and Bowie, J. U. (2005) Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14278-14283.
67. Plotkowski, M. L., Kim, S., Phillips, M. L., Partridge, A. W., Deber, C. M. and Bowie, J. U. (2007) Transmembrane domain of myelin protein zero can form dimers: possible implications for myelin construction. *Biochemistry* **46**, 12164-12173.
68. Barwe, S. P., Kim, S., Rajasekaran, S. A., Bowie, J. U. and Rajasekaran, A. K. (2007) Janus model of the Na,K-ATPase beta-subunit transmembrane domain: distinct faces mediate alpha/beta assembly and beta-beta homo-oligomerization. *J. Mol. Biol.* **365**, 706-714.
69. Reiersen, H. and Rees, A. R. (2001) The hunchback and its neighbours: proline as an environmental modulator. *Trends Biochem. Sci.* **26**, 679-684.
70. Bright, J. N., Shrivastava, I. H., Cordes, F. S. and Sansom, M. S. (2002) Conformational dynamics of helix S6 from Shaker potassium channel: simulation studies. *Biopolymers* **64**, 303-313.
71. Tieleman, D. P., Shrivastava, I. H., Ulmschneider, M. R. and Sansom, M. S. (2001) Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. *Proteins* **44**, 63-72.
72. Jin, T., Peng, L., Mirshahi, T., Rohacs, T., Chan, K. W., Sanchez, R. and Logothetis, D. E. (2002) The (beta) gamma subunits of G proteins gate a K(+) channel by pivoted bending of a transmembrane segment. *Mol. Cell.* **10**, 469-481.
73. Wigley, W. C., Corboy, M. J., Cutler, T. D., Thibodeau, P. H., Oldan, J., Lee, M. G., Rizo, J., Hunt, J. F. and Thomas, P.

J. (2002) A protein sequence that can encode native structure by disfavoring alternate conformations. *Nat. Struct. Biol.* **9**, 381-388.
74. Lasso, G., Antoniw, J. F. and Mullins, J. G. (2006) A combinatorial pattern discovery approach for the prediction of membrane dipping (re-entrant) loops. *Bioinformatics* **22**, e290-297.
75. Rapp, M., Granseth, E., Seppala, S. and von Heijne, G. (2006) Identification and evolution of dual-topology membrane proteins. *Nat. Struct. Mol. Biol.* **13**, 112-116.
76. Bowie, J. U. (2006) Flip-flopping membrane proteins. *Nat. Struct. Mol Biol.* **13**, 94-96.
77. Choi, S., Jeon, J., Yang, J. S. and Kim, S. (2008) Common occurrence of internal repeat symmetry in membrane proteins. *Proteins* **71**, 68-80.
78. Jeon, J., Yang, J. S. and Kim, S. (2009) Integration of evolutionary features for the identification of functionally important residues in Major Facilitator Superfamily (MFS) transporters. *PLoS Comput. Biol.* **5**, e1000522.
79. Tang, S., Liao, J. C., Dunn, A. R., Altman, R. B., Spudich, J. A. and Schmidt, J. P. (2007) Predicting allosteric communication in myosin via a pathway of conserved residues. *J. Mol. Biol.* **373**, 1361-1373.
80. Suel, G. M., Lockless, S. W., Wall, M. A. and Ranganathan, R. (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat. Struct. Biol.* **10**, 59-69.
81. Lao, D. M. and Shimizu, T. (2001) A method for discriminating a signal peptide and a putative 1st transmembrane segment. Proceedings of the 2001 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences - METMBS '01. pp. 119-125, CSREA Press, USA.
82. Lao, D. M., Arai, M., Ikeda, M. and Shimizu, T. (2002) The presence of signal peptide significantly affects transmembrane topology prediction. *Bioinformatics* **18**, 1562- 566.
83. Kall, L., Krogh, A. and Sonnhammer, E. L. (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027-1036.
84. Walz, T., Hirai, T., Murata, K., Heymann, J. B., Mitsuoka, K., Fujiyoshi, Y., Smith, B. L., Agre, P. and Engel, A. (1997) The three-dimensional structure of aquaporin-1. *Nature* **387**, 624-627.
85. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. and MacKinnon, R. (1998) The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* **280**, 69-77.
86. Granseth, E., Viklund, H. and Elofsson, A. (2006) ZPRED: predicting the distance to the membrane center for residues in alpha-helical membrane proteins. *Bioinformatics* **22**, e191-196.
87. von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic. Acids. Res.* **31**, 258-261.