Structure Prediction of Membrane Proteins

Chunlong Zhou[1], Yao Zheng[2], and Yan Zhou[1]*

[1] *Hangzhou Genomics Institute/James D. Watson Institute of Genome Sciences, Zhejiang University/Key Laboratory of Bioinformatics of Zhejiang Province, Hangzhou 310008, China;* [2] *Center for Engineering and Scientific Computation, Zhejiang University, Hangzhou 310027, China.*

There is a large gap between the number of membrane protein (MP) sequences and that of their decoded 3D structures, especially high-resolution structures, due to difficulties in crystal preparation of MPs. However, detailed knowledge of the 3D structure is required for the fundamental understanding of the function of an MP and the interactions between the protein and its inhibitors or activators. In this paper, some computational approaches that have been used to predict MP structures are discussed and compared.

Key words: structure prediction, membrane proteins

Introduction

Membrane proteins (MPs) constitute about 30% of all the proteins encoded in the currently known genomes, and play critical roles in cell signaling, ion transport, and cell-cell communications, as well as assist the folding of other MPs (*1*). Because of these biological significance, MPs represent the most important class of drug targets—about 50% of current molecular targets are membrane-bound (*2*). However, only about 2% (518 of 25,176) of the 3D structures deposited in the Protein Data Bank (PDB; ref. *3*) are for MPs. And the number of high-resolution structures (from X-ray diffraction and more recently from NMR) remains even smaller, largely because of the difficulties in crystallizing MPs. Recently, some new ideas and experimental approaches have been introduced in the area of MP crystallization (*4*), all of which exploit the spontaneous self-assembling properties of lipids and detergent as vesicles (vesicle-fusion method), discoidal micelles (bicelle method), and liquid crystals or mesophases (in meso or cubic-phase method). Despite these promising new methods, the current gap between need and supply of MP 3D structures makes prediction algorithms important and essential.

MPs come in a variety of sizes and shapes, though the available 3D structure principles are far less diverse than those of the globular proteins. From a structural point of view, there are two major groups of MPs. One is the α-helix bundle protein, in which one or several α-helices span the membrane; and the other is β-barrel protein, in which eight or more antiparallel TM β-strands form a closed barrel (*5*, *6*). Two recent examples (*7*, *8*) are shown in Figure 1.

Since Jähnig and Edholm in 1992 presented one of the first methods using secondary structure prediction to build suitable model structures as initial conformations for molecular dynamic studies (*9*), several groups have tried computational approaches to elucidate MP structures. In 1993, Milik and Skolnick presented a method based on the combination of a hydropathy scale for the prediction of trans-bilayer fragments with dynamic Monte Carlo simulation techniques (*10*). In 1994, Taylor *et al* adapted some programs originally developed for the prediction of globular protein structures to derive a method for the prediction of integral MP structures (*11*). Each step in the method is fully automated, from the initial sequence data bank searches to the final construction of 3D models. The major problem of MP prediction is lack of high-resolution experimental data. Consequently, estimates for prediction accuracy are perhaps overly optimistic. Here, we summarize recent attempts within the field of computational biology and bioinformatics to predict an MP's structure.

Secondary Structure Prediction and Transmembrane Segments Topology Prediction

Most current methods of theoretical MP structure prediction do not actually deal with predicting the

*** Corresponding author.**
E-mail: zhouyan@genomics.org.cn

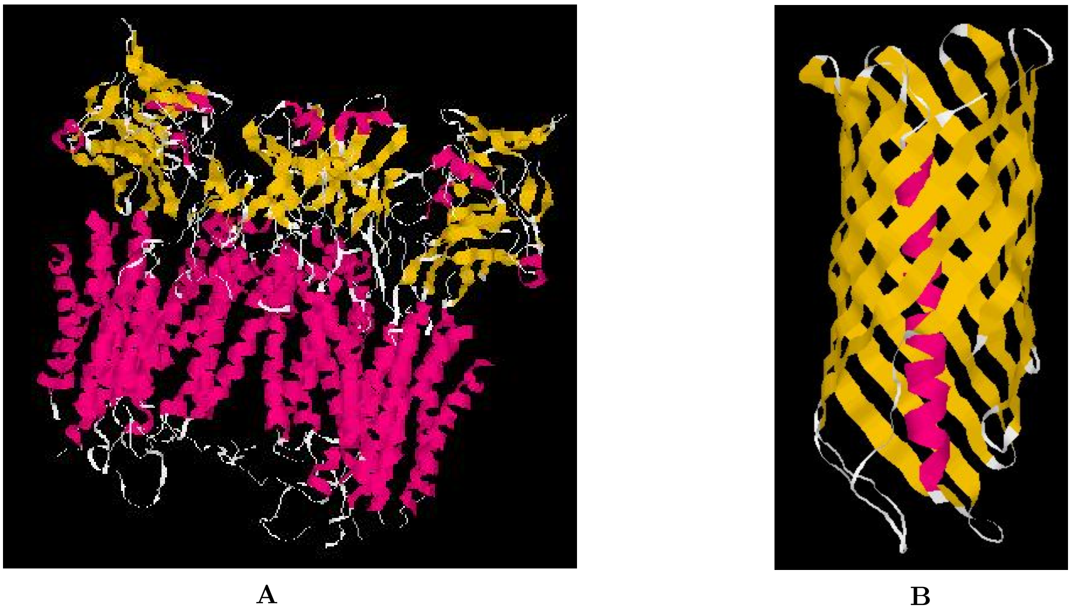

Fig. 1 The crystal structures of two new MPs by X-ray diffraction. **A.** The cytochrome B6F complex of an α-helix bundle protein from *Mastigocladus Laminosus*, PDB Id: 1VF5 (*7*). The red helices are the TM α-helix segments. **B.** The translocator domain of autotransporter nalp of a β-barrel protein from *Neisseria Meningitidis*, PDB Id: 1UYN (*8*). The yellow segment is the TM β-barrel composed of 12 membrane strands, and an N-terminal α-helix is in the center of the barrel.

3D structure, but rather try to predict the most likely topology of the protein, that is to say, the in/out location of the N and C termini relative to the membrane, and the number and position of transmembrane (TM) segments. A high-quality model of secondary structure and topology is a prerequisite for experimental structure-function studies, and can be a starting point for attempts to model the 3D structure before molecular dynamics or simulated annealing simulations. In recent years, various accurate methods have been applied to the topology prediction of TM α-helices and β-strands, respectively. Table 1 shows the main methods of TM segment topology prediction. Because the number of high-resolution structures of β-barrel proteins is less than that of the α-helix proteins, the neural network has been more frequently adopted in the β-strand topology prediction. The details of some methods based on Hidden Markov Models (HMMs) are listed in Table 2.

Many secondary structure prediction methods are based on statistical methods, physicochemical methods, sequence pattern maching, and evolutionary conservation (*12*). The main methods for identifing TM helices are on the basis of their hydrophobicity and known minimum length (at least 15 residues; ref. *13*). Membrane propensities were defined by a statistical analysis carried out on a set of 640 TM helices, belonging to 133 MPs extracted from SWISS-PROT (*14*) that have experimentally defined topologies.

The five widely used prediction methods for predicting the topology of α-helix bundle MPs are TMHMM (*15*), HMMTOP (*16*), MEMSAT (*17*), PHDhtm (*18*), and TopPred (*19*). TMHMM, HMMTOP, and MEMSAT are all based on HMMs with 5~7 types of structural states. PHDhtm is designed to use information from homologous proteins. TopPred was the first topology prediction method that combined hydrophobicity analysis and the positive-inside rule. Generally, these sequence-based methods for predicting the number and approximate location of TM helices within MPs have about 85% accuracy. In 2003, Karin Melén *et al* tried to construct useful reliability scores for these methods (*20*). They estimated an overall topology prediction accuracy of 55%-60% when entire proteomes are analyzed. The DAS (dense alignment surface; ref. *21*) algorithm can provide a solution to the problem that non-transmembrane query sequences may give false positive hits (20%-30%) in the prediction process. The upgraded and modified version of the DAS-prediction method, DAS-TMfilter algorithm, has been distributed (*22*). The new algorithm is designed to make distinction between protein sequences with and without TM helices at a reasonably low rate of false positive prediction (~1 among

Table 1 The Main Methods of Transmembrane Segments Topology Prediction

Segment type	Method	Approach	Self-proclaimed accuracy (segments)	Self-proclaimed accuracy (proteins)
Transmembrane α-helices	TMHMM	HMM	97%-98%	77%-78%
	HMMTOP	HMM	>98%	85%
	MEMSAT	HMM	92%	77%
	PHDhtm	homologous & neural network	98%	89%
	TopPred	hydrophobicity analysis & positive-inside rule	–	96%
	DAS-TMfilter	dense alignment surface	–	95%
	ConPred_elite	consensus approach	–	95%-98%
Membrane β-strands	Gromiha's	based on the conformational parameters and surrounding hydrophobicities	–	82%
	Diederichs's	neural network	–	–
	Jacoboni's	neural network	93%	78%
	Martelli's	HMM	–	84%

Table 2 Several Methods Based on Hidden Markov Model

Method	Number of states	Type of states
TMHMM	7	helix core, helix caps on either side of the membrane, short loop on cytoplasmic side/inside, short and long loop on noncytoplasmic side/outside, and a globular domain state
HMMTOP	5	inside loop, inside helix tail, helix, outside helix tail, and outside loop
MEMSAT	5	inside loop, inside helix tail, helix, outside helix tail, and outside loop
Martelli's	6	2 β-strand cores and 1 β-strand cap on either side of the membrane; 1 inner loop, 1 outer loop, and 1 globular domain state in the middle of each loop

100 unrelated queries) while the high efficiency of the original algorithm locating TM segments in queries is preserved (sensitivity of ~95% among documented proteins with helical TM regions). In 2003, Xia and colleagues presented a new approach, ConPred_elite (*23*), that can predict the whole topology with accuracies of 98% for prokaryotic and 95% for eukaryotic proteins as they reported.

Besides the TM helix, another TM segments type is β-barrel, which consists of several TM strands. Unlike α-helical MPs, there are no simple low-resolution experiments that yield large amounts of data for β-barrel MPs. This has constrained the ability to develop prediction methods. All early attempts to predict membrane strands employed the amphipacity and hydrophobicity of β-strands. Unfortunately, membrane strands have no long stretch of consecutive hydrophobic residues. In fact, the overall hydrophobicity for β-barrel MPs is similar to that of soluble proteins (*13*).

Gromiha and colleagues combined amino acid preferences for β-strands with the surrounding hydrophobicity of the respective residues to predict β-strands (*24*). They reproduced about 82% of the residues in structure-known membrane regions. Diederichs and colleagues proposed to use a neural network to predict the topology of the bacterial outer membrane β-strand proteins and to locate residues along the axes of the pores (*25*). Jacoboni and colleagues applied a method combining neural networks and dynamic programming to predict the location of membrane strands (*26*). The authors estimated that their system correctly predicted about 93% of all known membrane strands. More recently, Martelli *et al* developed a sequence-profile-based HMM model that can predict the topology of β-barrel MPs cycling with 6 types of states (*27*). They reported that the accuracy of per residue of the model was about 83%.

Lately the following protocol starting from secondary structure prediction and TM segments topology prediction are often used. Secondary structure prediction followed by TM segments identification along with prediction of loops connecting the segments, and molecular dynamics or simulated annealing simulations, may be finally used to refine these primal models. During the last refinement step, the protein is often inserted into a water/lipid bilayer/water or a water/n-octane/water environment to take into account the presence of the cell membrane. CHARMM, GROMOS, Amber, and cvff-insight are some widely used force fields in molecular dynamics calculation. The slow dynamics of lipid molecules in the bilayer might bring the difficulties in equilibrating the system (*28*).

The Direct Prediction of Whole 3D Structures

For globular proteins, the major successful methods for structure prediction include homology modeling, threading, and *ab initio* folding. Along with lucubrating the mechanism of MP folding and increasing the number of high-resolution MP structures, these methods will been applied to the direct prediction of whole MP 3D structures.

The question of how the controlled integration of an MP into the lipid bilayer takes place is still not fully worked out, and there are certainly aspects of MP structures that will probably not be fully appreciated until this step has been accomplished. Some pursuers educed the viewpoint that the prediction of MP structures from amino acid sequences was, in large measure, a problem of physicochemistry (*29*). Physical influences that shape MP structures include interactions of the polypeptide chains with water, bilayer hydrocarbon core, bilayer interfaces, and cofactors. Studies on the mechanism of insertion and folding of MPs into membranes are relatively rare and have been mostly performed with two model proteins: bacteriorhodopsin (BR; ref. *30*) of *Halobium salinarium* and outer MP A (OmpA; ref. *31*) of *Escherichia coli.* While BR is a representative α-helical bundle protein, OmpA belongs to the class of β-barrel protein.

Homology modeling constructs structures (targets) that are homologous to other protein(s) whose 3D structure is known (templates). It bases mainly on the conservation of protein folds rather than primary sequences homology. Because few high-resolution MP 3D structures are available to be used as templates, and the modeling can be unreliable when the sequence identity between the template and target proteins falls below 20%-30%, the applicability of homology modeling is limited. The same difficulties must been envisaged for threading methods.

In 2003, an *ab initio* method was presented (*32*), whose knowledge-based technique added a membrane potential to the energy terms (pairwise, solvation, steric, and hydrogen bonding). The method is based on the assembly of supersecondary structural fragments taken from a library of highly resolved protein structures using a standard simulated annealing algorithm. Results obtained by applying the method to small MPs of known 3D structures showed that the method is able to predict, at a reasonable accurate level, both the helix topology and the conformations of these proteins.

Conclusion

The structure prediction of membrane proteins still remains an interesting scientific problem. Because of the physical difference between MPs and GPs (globular proteins), more efforts have been put upon TM segment topology prediction for MPs. Current segment accuracy of reported algorithms are pretty high (above 90%), while the overall accuracy are still around 50%-60%, which gives birth to hand-raising methods to combine the reports from several other algorithms. The lack of both high-resolution and low-resolution experimental data of MP structures makes the algorithm development and their evaluation difficult, but the fact that most MP sequences are used as space blocks to get through the membrane bilayer that has predefined thickness makes the structure prediction of MPs simple on functional aspects. New algorithms will emerge and reported algorithms will be refined to give a better answer to this problem.

References

1. White, S.H. and Wimley, W.C. 1999. Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28: 319-365.
2. Drews, J. 2000. Drug discovery: a historical perspective. *Science* 287: 1960-1964.
3. Berman, H.M., *et al.* 2000. The protein data bank. *Nucleic Acids Res.* 28: 235-242.
4. Caffrey, M. 2003. Membrane protein crystallization. *J. Struct. Biol.* 142: 108-132.

5. Henderson, R. and Unwin, P.N. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* 257: 28-32.
6. Koebnik, R., *et al.* 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37: 239-253.
7. Kurisu, G., *et al.* 2003. Structure of the cytochrome B6F complex of oxygenic photosynthesis: tuning the cavity. *Science* 302: 1009-1014.
8. Oomen, C.J., *et al.* 2004. Structure of the translocator domain of a bacterial autotransporter. *EMBO J.* 23: 1257-1266.
9. Jähnig, F. and Edholm, O. 1992. Modeling of the structure of bacteriorhodopsin: a molecular dynamics study. *J. Mol. Biol.* 226: 837-850.
10. Milik, M. and Skolnick, J. 1993. Insertion of peptide chains into lipid membranes: an off-lattice Monte Carlo dynamics model. *Proteins* 15: 10-25.
11. Taylor, W.R, *et al.* 1994. A method for alpha-helical integral membrane protein fold prediction. *Proteins* 18: 281-294.
12. Rost, B. 2001. Protein secondary structure predication continues to rise. *J. Strct. Biol.* 134: 204-218.
13. Chen, C.P. and Rost, B. 2002. State-of-the-art in membrane protein prediction. *Appl. Bioinformatics* 1: 21-35.
14. Bairoch, A. and Apweiler, R. 1997. The SWISS-PROT protein sequence database: its relevance to human molecular medical research. *J. Mol. Med.* 75: 312-316.
15. Krogh, A., *et al.* 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305: 567-580.
16. 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.
17. Jones, D.T., *et al.* 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry* 33: 3038-3049.
18. Rost, B., *et al.* 1996. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci.* 5: 1704-1718.
19. von Heijne, G. 1992. Membrane protein structure prediction—hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 225: 487-494.
20. Karin, M., *et al.* 2003. Reliability measures for membrane protein topology prediction algorithms. *J. Mol. Biol.* 327: 735-744.
21. Cserzo, M., *et al.* 2002. On filtering false positive transmembrane protein predictions. *Protein Engin.* 15: 745-752.
22. Cserzo, M., *et al.* 2004. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. *Bioinformatics* 20: 136-137.
23. Xia, J.X., *et al.* 2004. ConPred_elite: a highly reliable approach to transmembrane topology prediction. *Comput. Biol. Chem.* 28: 51-60.
24. Gromiha, M.M, *et al.* 1997. Identification of membrane spanning beta strands in bacterial porins. *Protein Engin.* 10: 497-500.
25. Diederichs, K., *et al.* 1998. Prediction by a neural network of outer membrane beta-strand protein topology. *Protein Sci.* 7: 2413-2420.
26. Jacoboni, I., *et al.* 2001. Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor. *Protein Sci.* 10: 779-787.
27. Martelli, P.L., *et al.* 2002. A sequence-profile-based HMM for predicting and discriminating beta-barrel membrane proteins. *Bioinformatics* 18: 46-53.
28. Faraldo-Gomez, J.D, *et al.* 2002. Setting up and optimization of membrane protein simulations. *Eur. Biophys. J.* 31: 217-227.
29. White, S.H. 2003. Translocons, thermodynamics, and the folding of membrane proteins. *FEBS Letters* 555: 116-121.
30. Booth, P.J. and Curran, A.R. 1999. Membrane protein folding. *Curr. Opin. Struct. Biol.* 9: 115-121.
31. Kleinschmidt, J.H., *et al.* 1999. Outer membrane protein A of *E. coli* inserts and folds into lipid bilayers by a concerted mechanism. *Biochemistry* 38: 5006-5016.
32. Pellegrini-Calace, M., *et al.* 2003. Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures. *Proteins* 50: 537-545.