Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Topology of membrane proteins — predictions, limitations and variations

Konstantinos D .Tsirigos[5], Sudha Govindarajan[1,2], Claudio Bassot[1,2], Åke Västermark[1,2,7], John Lamb[1,2], Nanjiang Shu[1,2,3,4] and Arne Elofsson[1,2,6]

Transmembrane proteins perform a variety of important biological functions necessary for the survival and growth of the cells. Membrane proteins are built up by transmembrane segments that span the lipid bilayer. The segments can either be in the form of hydrophobic alpha-helices or beta-sheets which create a barrel. A fundamental aspect of the structure of transmembrane proteins is the membrane topology, that is, the number of transmembrane segments, their position in the protein sequence and their orientation in the membrane. Along these lines, many predictive algorithms for the prediction of the topology of alpha-helical and beta-barrel transmembrane proteins exist. The newest algorithms obtain an accuracy close to 80% both for alpha-helical and beta-barrel transmembrane proteins. However, lately it has been shown that the simplified picture presented when describing a protein family by its topology is limited. To demonstrate this, we highlight examples where the topology is either not conserved in a protein superfamily or where the structure cannot be described solely by the topology of a protein. The prediction of these non-standard features from sequence alone was not successful until the recent revolutionary progress in 3D-structure prediction of proteins.

Addresses
[1] Science for Life Laboratory, Stockholm University, SE-171 21 Solna, Sweden
[2] Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden
[3] National Bioinformatics Infrastructure, Sweden
[4] Nordic e-Infrastructure Collaboration, Norway
[5] Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany
[6] Swedish e-Science Research Center (SeRC), Sweden
[7] NITECH, Showa-Ku, Nagoya 466-8555 Japan

Corresponding author: Elofsson, Arne (arne@bioinfo.se)

Current Opinion in Structural Biology 2018, **50**:9–17

This review comes from a themed issue on **Sequences and topology**

Edited by **Joseph A Marsh and Sarah A Teichmann**

For a complete overview see the Issue and the Editorial

Available online 5th November 2017

http://dx.doi.org/10.1016/j.sbi.2017.10.003

Introduction

Transmembrane (TM) proteins are the class of membrane proteins which cross the lipid bilayer. They can be broadly classified into two structural categories; the ones that span the membrane in the form of alpha-helices and the ones whose TM regions are composed of beta-strands in the form of anti-parallel barrels.

Alpha-helical TM proteins constitute the most important and the most widely studied category of membrane proteins. They typically comprise 25–30% of all proteins encoded in a genome [1] and carry out a series of functions crucial to the life of the cells. These include cellular recognition, molecular receptors, passive and active transport of substances via the membrane, signal transduction, protein secretion and enzymatic activity [2].

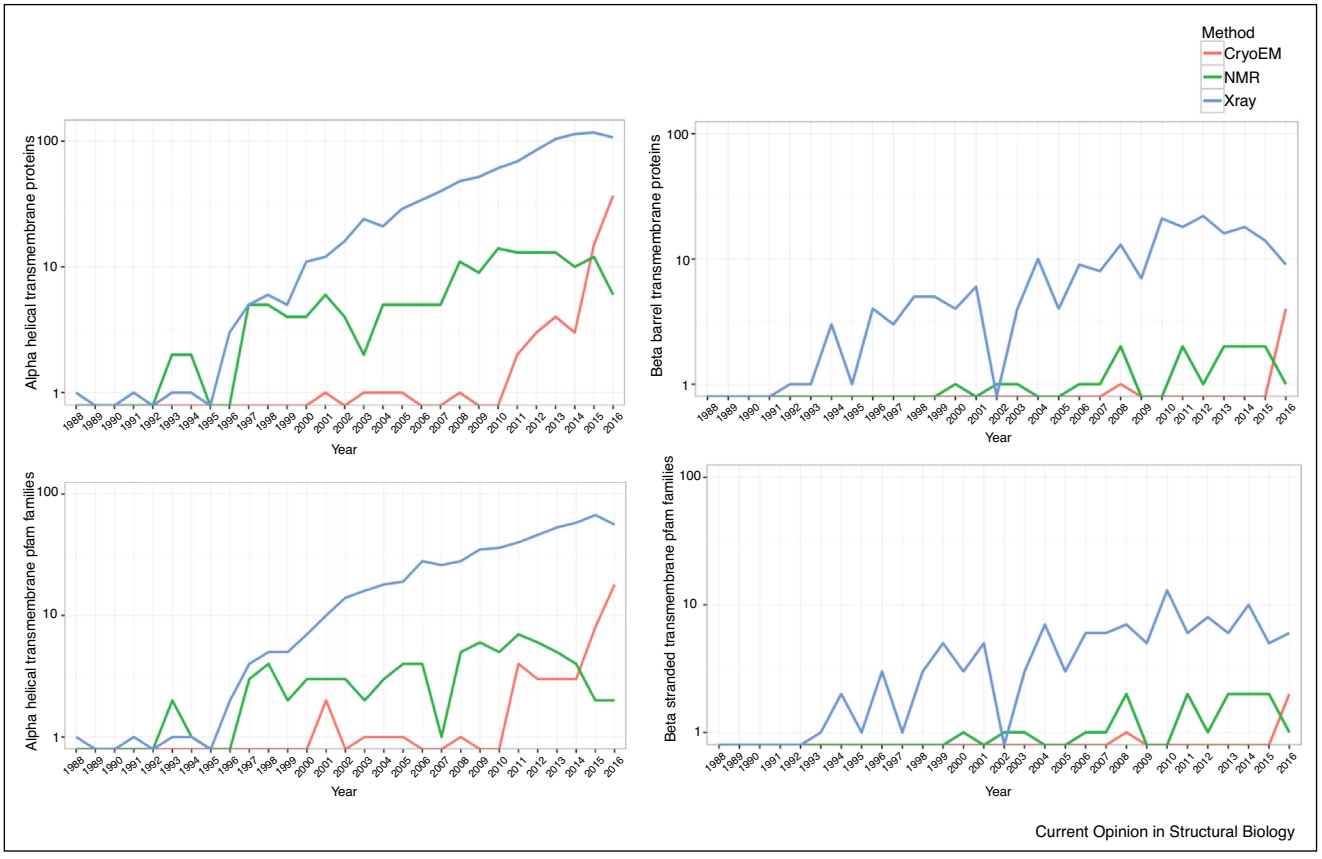
A beta-barrel can be defined as a beta-sheet that coils and loops forming a closed structure in the shape of a barrel. TM beta-barrel proteins are further divided into several groups, mainly based on their structural similarity, which, in most cases, reflects functional similarities as well. In contrast to alpha-helical membrane proteins that are abundant in virtually all cellular membranes, beta-barrels only exist in the outer membranes of Gram-negative bacteria and in chloroplasts and mitochondria [3].

Transmembrane protein structures

The number of solved membrane protein structures has been steadily increasing for two decades (Figure 1). The vast majority of deposited 3D-structures in PDB [4] are derived from X-ray crystallography, but, lately, the number of structures solved by cryo-electron (cryo-EM) microscopy is increasing [5]. If this trend continues, more structures will be solved by cryo-EM than by X-ray in the coming years; given the difficulty to crystallise some membrane proteins, this will most likely unravel even more structural diversity of TM proteins.

In Figure 1, alpha-helical and beta-stranded TM structures that are associated with Pfam [6] families are shown for every year. Even though a lot of alpha-helical TM structures have been determined so far, these are

Figure 1



The total number of alpha helical and beta-stranded membrane protein structures solved per year (Log scale) by X-ray, NMR and CryoEM methods are shown. The total number of alpha helical and beta-stranded membrane protein structures (Log scale) that could be associated with Pfam families per year by Xray, NMR and CryoEM methods are shown.It can be seen that since 2016 the structures solved by X-ray crystallography has decreased, but the structures solved by CryoEM is increasing.

associated with only 171 Pfam families (Figure 1). However, topology predictions by our recently published method, SCAMPI2 [7], show 1059 Pfam families consisting of TM domains. A lot of TM structures are yet to be solved to fill in the membrane protein space.

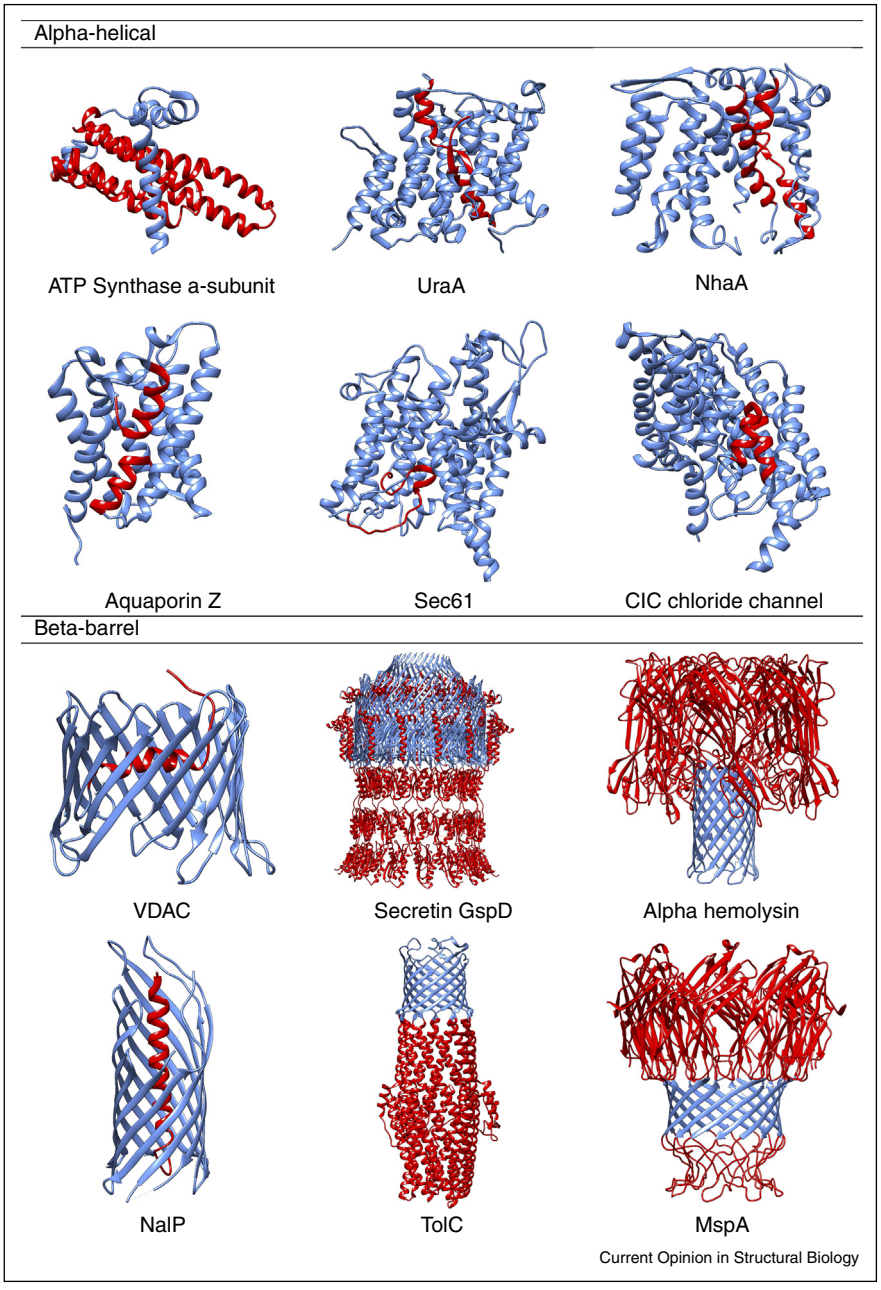
Topologies and variations observed in transmembrane proteins

For a long time, it was believed that alpha-helical TM proteins were simply bundles of straight helices. Today, with more than 1000 structures solved, it is clear that this is not the case. The structural repertoire of membrane proteins is much more complex than what was believed only a few years ago [8,9]. Non-canonical structural features, such as interface helices [10], re-entrant regions [8] and deep core-coil residues [9] are frequent in TM proteins.

A number of variations are found in channels or transporters, and, usually, play a functional role. Some examples of non-standard topologies are shown in Figure 2, starting from the recently solved horizontal TM helices topology of the a-subunit of ATP synthase [11••,12]. This peculiar structure shows a membrane-intrinsic alpha-helix tilted by 70° relative to close standard perpendicular TM helices. In the ATP synthase complex, the horizontal helices line two aqueous half channels in the inner mitochondrial membrane essential for proton translocation [11••]. Another class of non-standard topology, termed the discontinuous ('spiny') helices, is found in ion transporters like NhaA [13] or secondary carriers like UraA [14•,15]. These regions seem to be essential for the binding of the substrates and may confer flexibility to the structures, allowing thus the conformational change of the transporter [16].

Further, the re-entrant regions are essentially membrane-penetrating regions that enter and exit the membrane on the same side. Examples of functionally important re-entrant regions are: firstly, Aquaporin Z, in which the two re-entrant coil-helix domains form the selectivity filter [17]; secondly, the Sec61 protein-conducting channel,

Figure 2



PDB structures exemplifying topological variation. The non-standard domains are highlighted in red. In the alpha helical panel are shown: ATP synthase a-subunit (5ARA), UraA (3QE7) and NhaA (1ZCD) (regions 86–100 and 260–281 in UraA and 94–116 and 152–164 in NhaA are hidden for reason of clarity). Aquaporin Z (1RC2), Sec61 (1RH5) (residues 50–66 of chain B are hidden) and ClC chloride channel (1OTS). In the beta-barrel panel, the structures of Voltage Dependent Anion Channel (VDAC) (2JK4), Secretin GspD (5WQ8), Alpha hemolysin (7AHL), NalP(1UYN), TolC (1EK9), MspA (1UUN) are shown.

where a re-entrant coil-helix-coil domain regulates the permeability of the translocation pore [18]; and finally the ClC chloride channel, showing conserved re-entrant helix–coil–helix domains [19]. An overview of these examples can be seen in Figure 2.

Internal symmetry in alpha-helical transmembrane proteins

One striking difference between soluble and membrane proteins is that soluble multi-domain proteins are very common in higher organisms [20], while the combination

of membrane domains seems to be rare [21]. It has been argued that the membrane puts constrains on the protein such that domain fusion is very unlikely to occur [21] and that non-covalent oligomeric associations, which are common in membrane proteins, may provide an alternative source of evolutionary diversity.

In contrast to the rareness of domain fusion, internal symmetry due to the fusion of two homologous genes is found in about 50% of the larger alpha-helical membrane proteins [22] and it has been proposed that all beta-hairpins in beta-barrel proteins have a common origin [23]. In both these cases, it is clear that the symmetry offers such an advantage that basically all traces of the original smaller proteins have been lost through evolution. The symmetry is often related to the mobile function of the protein, where conformational transitions are mirrored between the halves and internal duplication has mechanistic significance [24].

Evolution of repeats has been studied in solute carrier MFS transporters. Rearrangements of the triple helical repeat in two families of this clan, namely fucose permease and lactose permease (LacY) show high similarity [25•,26•]. This evolutionary mix-and-match of transporters shows how the primordial helical-triplets have been assembled in different order to give rise to diversity in MFS transporters.

Variation in topology between homologous proteins

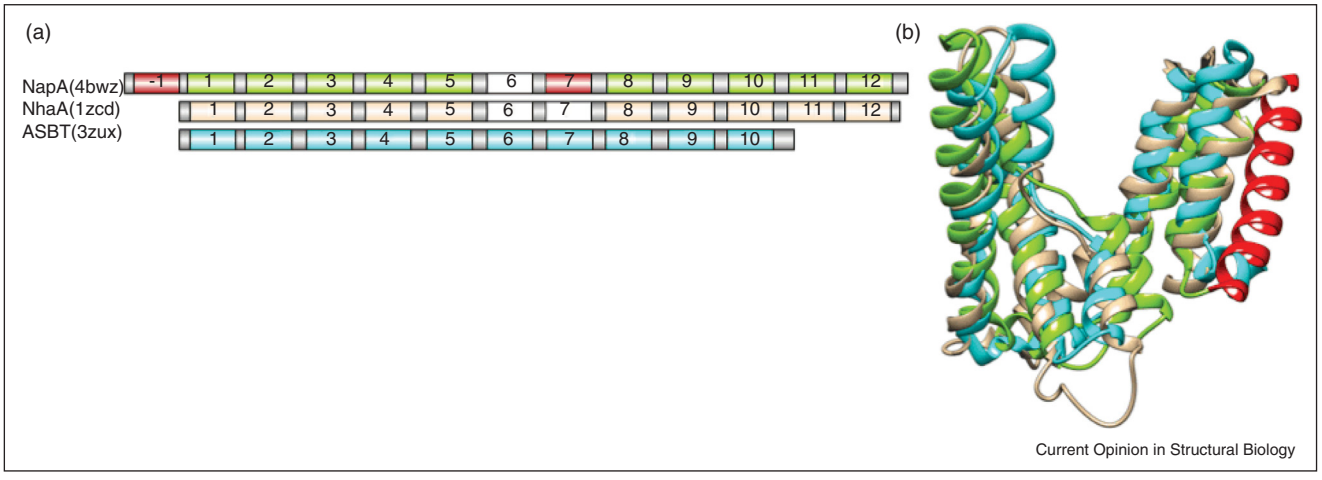
Alterations in topology have been recently observed in the cytochrome P450 enzyme [27•]. Paralogs exist with one and two TM helices. Both the paralogs show cinnamate hydroxylase activity. The paralog with the single-TM helix targets the protein to the ER with (N) lumen-(C) cytosol orientation, while, the two-TM one shows dual topology with two locations of catalytic domain either in the cytosol or ER lumen.

Another interesting example of alterations in membrane topology can be observed in the monovalent cation:proton antiporter (CPA)/anion transporter (AT) clan (Pfam clan: CL0064) (Figure 3). These are solute carrier transporters that mediate the flow of various substances through the cell membrane. The internal symmetry in these proteins plays a vital role for conformational changes required for the transport process. Structural information is available for three out of 13 Pfam families in this clan, namely the 13-TM protein NapA from *Thermus thermophilus* (PF00999) [28], the 12-TM protein NhaA from *Escherichia coli* (PF06965) [13] and the 10-TM protein ASBT from *Neisseria meningitidis* (PF01758) [29]. All these proteins contain inverted repeats of five or six TM helices. Structural superposition of the repeats from NhaA, NapA and ASBT shows the structural similarity of the repeated unit (Figure 3).

Beta-barrel proteins

Beta-barrel proteins are involved in many biological processes. The smallest beta-barrels, OmpA [30] and OmpX [31] contain 8 TM beta-strands. In most cases, the number of beta-strands ranges from 8 to 24 [32]; for example the long-chain fatty acids transporters, FadL, has 14 beta-strands [33] and the nucleoside transporter (Tsx) contains 12 beta-strands [34]. Porins, such as OmpF [35] and PhoE [36] consist of 16 beta-strands and they mediate passive

Figure 3



Variations of topology in CPA transporter superfamily: the CPA transporter superfamily consists of proteins NapA, NhaA and ASBT belonging to different Pfam families. NapaA, NhaA and ASBT are composed of 13, 12 and 10 TM helices respectively. TM helices involved in repeats are coloured in green (NapA), light brown (NhaA) and cyan (ASBT). Helices not involved in the repeats are coloured white. NhaA and ASBT have repeats with five TM helices. NapA has repeats that are six TM helices long. The extra helix involved in the repeats is shown in red. Structural superposition of the repeats from the three proteins shows the structural similarity between the repeats.

transport of small molecules (e.g. ions) or water through the membrane. However, even larger beta-barrels are found, such as for the Usher protein PapC, which contains a beta-barrel with 26 beta-strands [37], the largest number reported so far for a single-barrel outer membrane protein.

Many of the aforementioned proteins have been experimentally shown to function as monomers, but there are also some where oligomerization is necessary in order to obtain the proper functionality. In the latter case we find the bacterial porins, which function as homotrimers [38] and beta-barrels with enzymatic activity, like OmpT and OMPLA, which are homodimers [39,40].

Non-typical cases of beta-barrel proteins

The structure of the human mitochondrial porin voltage dependent anion channel (VDAC) [41] revealed the first odd-numbered beta-barrel protein, since it has 19 beta-strands. It was also found to contain an alpha-helix located horizontally midway within the pore of the structure, see Figure 2.

Autotransporters, an essential part of the type-V bacterial secretion pathway, are related to beta-barrel proteins, as their C-termini form a TM-beta barrel formed pore in the outer membrane. The mature protein will then be transferred through this pore. An example of such proteins is NalP that contains 12 beta-strands [42].

Besides the typical single-chain beta-barrel proteins, there exist cases where beta strands contributed by more than one polypeptide chains form the barrel. One such example is the TolC protein [43] and its homologs in the Outer Membrane Factors family (Figure 2). TolC, which belongs to the tripartite drug efflux pumps protein system machinery, contains both alpha-helices and beta-sheets and crosses both the outer membrane and the periplasmic space. Three monomers of this protein form a continuous channel and each monomer contributes four beta-strands to the 12-stranded beta-barrel.

Alpha-haemolysin from the Gram-positive bacterium *Staphylococcus aureus* is active as a heptamer, where the TM region is formed by a 14-stranded beta-barrel, two of which are contributed by each of the seven monomers [44]. Furthermore, the structure of the outer membrane channel of the Gram-positive bacterium *Mycobacterium smegmatis* revealed some special characteristics; it is an octamer that creates a 'double barrel', where the lower part is a 16-stranded beta-barrel (two beta-strands contributed from each monomer) and the upper part extends and creates a second beta-barrel, again with 16 strands [45].

Very recently, the structure of the first secretin protein (GspD) was determined, where it was shown that the secretin domain constitutes a novel double beta-barrel channel, with at least 60 beta-strands in each barrel [46•].

Transmembrane protein structure prediction methods

Nowadays, we still lack structural representation for most TM protein families. Therefore, the need for computational tools that will predict the structure of a TM protein with high accuracy is imperative. Although rapid progress has occurred in the 3D-prediction of membrane proteins [47••,48••,49••], for large-scale analysis we still have to rely on topology prediction algorithms. Luckily, these have been improved in recent years.

Topology prediction of alpha-helical transmembrane proteins

All topology prediction methods of alpha-helical membrane proteins are based on the rules that govern the biogenesis of these proteins, that is, the insertion of hydrophobic segments into the membrane by the translocon and the orientation preference determined by the positive-inside rule [50]. The positive-inside rule was first implemented in the TopPred algorithm [51]. Later, in MEMSAT [52], a dynamic programming method to identify the optimal topology was introduced. The power of this simple methodology can be seen in SCAMPI [53••] which in our benchmarks [54] is the best method using only a single sequence. SCAMPI is very similar to the original MEMSAT method, but it uses an more accurate hydrophobicity scale [53••,55]. Later, Hidden Markov models (HMMs) were introduced in TMHMM [56] and HMMTOP [57]. In comparison to earlier approaches, which used a fixed hydrophobicity scale, HMMs had the advantage that, in theory, the optimal scale could be learnt from the training data. Gradually, more and more HMM-based methods made use of evolutionary information, in the form of Multiple Sequence Alignments (MSAs) [58].

In PHD [59], the average hydrophobicity of a segment was replaced by an Artificial Neural Network (ANN) to predict the probability of a segment to be part of a TM region. Later methods, including MEMSAT3 [60] and OCTOPUS [61], extended the idea of ANNs by combining them with a dynamic programming module in order to produce the optimal topology. The advantage is that the ANN also can take into account correlation within a window in the sequence.

Consensus-based methods, like TOPCONS [62] and CCTOP [63•], that combine the outputs from several predictors and create a consensus prediction using dynamic programming have also been presented (Table 1). TOPCONS2 [64••] is in our benchmarks the best-performing one for topology prediction and discrimination of alpha-helical TM proteins.

Recent studies conclude that even the best topology prediction methods reach an upper limit of ~80% overall accuracy, probably owing to the limited amount of experimental structures, sequencing/annotation errors and

Table 1

List of topology prediction servers

Server name	URL
Alpha-helical	
SCAMPI2	http://scampi.bioinfo.se/
MEMSAT3	http://bioinf.cs.ucl.ac.uk/?id=756
OCTOPUS	[part of the TOPCONS2 suite]
TMHMM	http://www.cbs.dtu.dk/services/TMHMM/
HMMTOP	http://www.enzim.hu/hmmtop/
CCTOP	http://cctop.enzim.ttk.mta.hu/
Alpha-helical signal peptides	
TOPCONS2	http://topcons.net/
Philius	http://www.yeastrc.org/philius/pages/philius/runPhilius.jsp
Phobius	http://phobius.sbc.su.se/
PolyPhobius	http://phobius.sbc.su.se/poly.html
SPOCTOPUS	[part of the TOPCONS2 suite]
MEMSAT-SVM	http://bioinf.cs.ucl.ac.uk/psipred/
Beta barrels	
BOCTOPUS2	http://boctopus.bioinfo.se/
PRED-TMBB2	http://www.compgen.org/tools/PRED-TMBB2/
PROFtmb	https://www.predictprotein.org/
BetAware	http://betaware.biocomp.unibo.it/BetAware
TMBETAPRED-RBF	https://www.rbf.bioinfo.tw/~sachen/BARRELpredict/TMBETAPRED-RBF.php
ConBBPRED	http://www.bioinformatics.biol.uoa.gr/ConBBPRED/

unusual sequence features, like re-entrant regions [54,64••].

It is estimated that ~10% of all TM proteins encoded in a genome contain re-entrant regions [8]. However, very few methods attempted to predict these and the ones that do are not very successful [61]. The problem of correct prediction of re-entrant helices arises because there are too few structures to properly train a prediction method and they are also rather different from each other.

One of the major problems of topology predictions is that signal peptides are erroneously predicted as TM segments because of their high hydrophobicity [65]. To tackle this problem, methods that simultaneously predict the topology of a protein and the presence of a signal peptide, like Phobius [66], PolyPhobius [67], and SPOCTOPUS [68] and Philius [69] were developed. TOPCONS2 [64••] nowadays offers improved predictions for all types of proteins in a proteome, including the ones that contain a signal peptide in their sequence.

Prediction of beta-barrel transmembrane membrane proteins

A variety of topology prediction methods are available for beta-barrels (Table 1). Initially, they were based on hydrophobicity analysis, like Beta-Barrel Finder [70] and, later on, statistical analyses like BOMP [71].

ANNs were the first machine-learning methods to be employed for the task of beta-barrel topology prediction; however, these methods did not prove to be very successful.

Initial examples of HMM-based methodologies include PRED-TMBB [72] and ProfTMB [73]. Recently, BOCTOPUS [74] was introduced. It is a hybrid SVM-HMM method that improved the topology prediction significantly. BOCTOPUS2 [75] and PRED-TMBB2 are the newest methodologies that further improve beta-barrel topology predictions. They both exploit the 'dyad-repeat' pattern that the beta-strands exhibit (lipid-facing and pore-facing residues). They perform on par with each other regarding topology prediction when tested on a non-redundant dataset of beta-barrel structures and clearly outperform all previously published methods [76•].

The only consensus method for beta-barrel topology prediction, to date, is ConBBPRED [77], which uses a dynamic programming algorithm to combine the results from different methods into a final prediction. However, the performance of this method does not surpass the performance of state-of-the art methods.

There are also methods that aim specifically at the discrimination of beta-barrels from other classes of proteins in proteome-wide analyses. The best one is HHomp [78•] but it is rather slow (since it relies on MSAs) and thus not ideal for proteome-wide analyses. PRED-TMBB2, which can also operate on single-sequence mode, performs on par with HHomp but it is orders of magnitude faster [76•].

3D prediction of transmembrane proteins

Given the limited success in prediction of re-entrant regions and other irregular structures of alpha-helical and beta-barrel membrane proteins, one can ask what the value of these methods are. There has recently been a revolution in structure prediction for both individual proteins and complexes. The basis for this is the development of contact prediction methods using direct coupling information [79]. In combination with the topology prediction methods described above, the predicted contacts can be used to predict the structure of individual soluble proteins [80,81], alpha-helical membrane proteins [47••,48••] and beta-barrels [49••]. Even some of the irregular topological elements, such as re-entrant regions and plug-domains could be modeled with some accuracy using these methods. The recent progress in contact predictions [82,83] has already enabled the predictions of hundred of TM protein families [84,85].

Conclusion

In this review, we summarise our knowledge regarding TM protein topology and related prediction methods. We highlight the fact that the simple topology prediction is

becoming more limited as more and more structures with non-standard topologies are discovered. We report on the recent progress made in predictions of alpha-helical membrane proteins [64••] and beta-barrel proteins [75•,76•]. In closing of this section, we note that the recent progress in 3D-structure predictions of membrane proteins will most likely provide structural insights into hundreds of membrane proteins and in the future it is likely that these methods will provide valuable structural insights into larger transmembrane complexes.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

This review is adapted from portions of Konstantinos Tsirigos PhD thesis. Arne Elofsson is supported by grants from the Wallenberg Foundation and Carl Tryggers Foundation and Åke Västermark is supported by the JSPS fellowship (PE16042).

This work was supported by grants from the Swedish Research Council (VR-NT 2009-5072, 2012-5046, VR-M 2010-3555), SSF, the Foundation for Strategic Research and Swedish E-science research center.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. 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.

2. Elofsson A, von Heijne G: **Membrane protein structure: prediction versus reality**. *Annu Rev Biochem* 2007, **76**:125-140.

3. Schulz G: **Transmembrane beta-barrel proteins**. *Adv Protein Chem* 2003, **63**:47-70.

4. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P: **The protein data bank**. *Nucleic Acids Res* 2000, **28**:235-242.

5. Earl L, Falconieri V, Milne J, Subramaniam S: **Cryo-EM: beyond the microscope**. *Curr Opin Struct Biol* 2017, **46**:71-78.

6. Finn R, Coggill P, Eberhardt R, Eddy S, Mistry J, Mitchell A, Potter S, Punta M, Qureshi M, Sangrador-Vegas A, Salazar G, Tate J, Bateman A: **The Pfam protein families database: towards a more sustainable future**. *Nucleic Acids Res* 2016, **44**: D279-85.

7. Peters C, Tsirigos K, Shu N, Elofsson A: **Improved topology prediction using the terminal hydrophobic helices rule**. *Bioinformatics* 2016, **32**:1158-1162.

8. 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.

9. Kauko A, Illergard K, Elofsson A: **Coils in the membrane core are conserved and functionally important**. *J Mol Biol* 2008, **380**: 170-180.

10. 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.

11. •• Allegretti M, Klusch N, Mills D, Vonck J, Kuhlbrandt W, Davies K: **Horizontal membrane-intrinsic alpha-helices in the stator a-subunit of an F-type ATP synthase**. *Nature* 2015, **521**:237-240.
The structure of the F-type ATP synthase contains a horizontal alpha-helix.

12. Zhou A, Rohou A, Schep D, Bason J, Montgomery M, Walker J, Grigorieff N, Rubinstein J: **Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM**. *Elife* 2015, **4**:e10180.

13. Hunte C, Screpanti E, Venturi M, Rimon A, Padan E, Michel H: **Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH**. *Nature* 2005, **435**:1197-1202.

14. • Lu F, Li S, Jiang Y, Jiang J, Fan H, Lu G, Deng D, Dang S, Zhang X, Wang J, Yan N: **Structure and mechanism of the uracil transporter UraA**. *Nature* 2011, **472**:243-246.
UraA contains a novel fold with two inverted repeats.

15. Västermark Å, Saier M Jr: **Evolutionary relationship between 5 +5 and 7+7 inverted repeat folds within the amino acid-polyamine-organocation superfamily**. *Proteins* 2014, **82**: 336-346.

16. Screpanti E, Hunte C: **Discontinuous membrane helices in transport proteins and their correlation with function**. *J Struct Biol* 2007, **159**:261-267.

17. Savage D, Egea P, Robles-Colmenares Y, O'Connell J, 3rd R, Stroud: **Architecture and selectivity in aquaporins: 2.5 Å X-ray structure of aquaporin Z**. *PLoS Biol* 2003, **1**:E72.

18. Van den Berg B, Clemons W Jr, Collinson I, Modis Y, Hartmann E, Harrison S, Rapoport T: **X-ray structure of a protein-conducting channel**. *Nature* 2004, **427**:36-44.

19. Dutzler R, Campbell E, MacKinnon R: **Gating the selectivity filter in CLC chloride channels**. *Science* 2003, **300**:108-112.

20. Ekman D, Elofsson A: **Identifying and quantifying orphan protein sequences in fungi**. *J Mol Biol* 2010, **396**:396-405.

21. Liu Y, Gerstein M, Engelman D: **Transmembrane protein domains rarely use covalent domain recombination as an evolutionary mechanism**. *Proc Natl Acad Sci U S A* 2004, **101**:3495-3497.

22. Hennerdal A, Falk J, Lindahl E, Elofsson A: **Internal duplications in alpha-helical membrane protein topologies are common but the nonduplicated forms are rare**. *Protein Sci* 2010, **19**:2305-2318.

23. Soding J, Remmert M, Biegert A: **HHrep: de novo protein repeat detection and the origin of TIM barrels**. *Nucleic Acids Res* 2006, **34(Web Server issue)**:W137-42.

24. McCoy J, Ren Z, Stanevich V, Lee J, Mitra S, Levin E, Poget S, Quick M, Im W, Zhou M: **The structure of a sugar transporter of the glucose EIIC superfamily provides insight into the elevator mechanism of membrane transport**. *Structure* 2016, **24**: 956-964.

25. • Madej M, Dang S, Yan N, Kaback H: **Evolutionary mix-and-match with MFS transporters**. *Proc Natl Acad Sci U S A* 2013, **110**:5870-5874.
The major facilitator superfamily is a good example of how topology can vary within a superfamily.

26. • Väastermark A, Saier M: **Major facilitator superfamily (MFS) evolved without 3-transmembrane segment unit rearrangements**. *Proc Natl Acad Sci U S A* 2014, **111**:E1162-3.
The MFS family has evolved through internal repeat expansions.

27. • Renault H, De Marothy M, Jonasson G, Lara P, Nelson D, Nilsson I, Andre F, von Heijne G, Werck-Reichhart D: **Gene duplication leads to altered membrane topology of a cytochrome p450 enzyme in seed plants**. *Mol Biol Evol* 2017, **34**:2041-2056.
One of the first examples of paralogous proteins with experimentally verified variation of topology.

28. Lee C, Kang H, von Ballmoos C, Newstead S, Uzdavinys P, Dotson D, Iwata S, Beckstein O, Cameron A, Drew D: **A two-domain elevator mechanism for sodium/proton antiport**. *Nature* 2013, **501**:573-577.

29. Hu N, Iwata S, Cameron A, Drew D: **Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT**. *Nature* 2011, **478**:408-411.

30. Morona R, Kramer C, Henning U: **Bacteriophage receptor area of outer membrane protein OmpA of *Escherichia coli* K-12**. *J Bacteriol* 1985, **164**:539-543.

31. Vogt J, Schulz GE: **The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence**. *Structure* 1999, **7**:1301-1309.

32. Fairman J, Noinaj N, Buchanan S: **The structural biology of beta-barrel membrane proteins: a summary of recent reports**. *Curr Opin Struct Biol* 2011, **21**:523-531.

33. van den Berg B, Black PN, Clemons WM, Rapoport TA: **Crystal structure of the long-chain fatty acid transporter FadL**. *Science* 2004, **304**:1506-1509.

34. Ye J, van den Berg B: **Crystal structure of the bacterial nucleoside transporter Tsx**. *EMBO J* 2004, **23**:3187-3195.

35. Danelon C, Suenaga A, Winterhalter M, Yamato I: **Molecular origin of the cation selectivity in OmpF porin: single channel conductances vs. free energy calculation**. *Biophys Chem* 2003, **104**:591-603.

36. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP: **Crystal structures explain functional properties of two *E. coli* porins**. *Nature* 1992, **358**:727-733.

37. Remaut H, Tang C, Henderson NS, Pinkner JS, Wang T, Hultgren SJ, Thanassi DG, Waksman G, Li H: **Fiber formation across the bacterial outer membrane by the chaperone/usher pathway**. *Cell* 2008, **133**:640-652.

38. Tamm LK, Arora A, Kleinschmidt JH: **Structure and assembly of beta-barrel membrane proteins**. *J Biol Chem* 2001, **276**: 32399-32402.

39. Dekker N, Tommassen J, Lustig A, Rosenbusch JP, Verheij HM: **Dimerization regulates the enzymatic activity of *Escherichia coli* outer membrane phospholipase A**. *J Biol Chem* 1997, **272**:3179-3184.

40. Kingma RL, Egmond MR: **Activation of a covalent outer membrane phospholipase A dimer**. *Eur J Biochem* 2002, **269**:2178-2185.

41. Bayrhuber M, Meins T, Habeck M, Becker S, Giller K, Villinger S, Vonrhein C, Griesinger C, Zweckstetter M, Zeth K: **Structure of the human voltage-dependent anion channel**. *Proc Natl Acad Sci U S A* 2008, **105**:15370-15375.

42. Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J, Gros P: **Structure of the translocator domain of a bacterial autotransporter**. *EMBO J* 2004, **23**:1257-1266.

43. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C: **Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export**. *Nature* 2000, **405**:914-919.

44. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE: **Structure of staphylococcal alpha-hemolysin a heptameric transmembrane pore**. *Science* 1996, **274**:1859-1866.

45. Faller M, Niederweis M, Schulz GE: **The structure of a mycobacterial outer-membrane channel**. *Science* 2004, **303**:1189-1192.

46. • Yan Z, Yin M, Xu D, Zhu Y, Li X: **Structural insights into the secretin translocation channel in the type II secretion system**. *Nat Struct Mol Biol* 2017, **24**:177-183.
The type II secretin system is an example of non-standard topology.

47. •• Nugent T, Jones D: **Accurate de novo structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis**. *Proc Natl Acad Sci U S A* 2012, **109**:E1540-7.
Accurate structural prediction of alpha-helical membrane proteins.

48. •• Hopf T, Colwell L, Sheridan R, Rost B, Sander C, Marks D: **Three-dimensional structures of membrane proteins from genomic sequencing**. *Cell* 2012, **149**:1607-1621.
Accurate structural prediction of alpha-helical membrane proteins.

49. •• Hayat S, Sander C, Marks D, Elofsson A: **All-atom 3d structure prediction of transmembrane beta-barrel proteins from sequences**. *Proc Natl Acad Sci U S A* 2015, **112**:5413-5418.
Accurate structural prediction of beta-barrel membrane proteins.

50. Heijne G: **The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology**. *EMBO J* 1986, **5**:3021-3027.

51. Claros M, von Heijne G: **TopPred II: an improved software for membrane protein structure predictions**. *Comput Appl Biosci* 1994, **10**:685-686.

52. Jones D, Taylor W, Thornton J: **A model recognition approach to the prediction of all-helical membrane protein structure and topology**. *Biochemistry* 1994, **33**:3038-3049.

53. •• 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.
This paper shows that basic understanding of the hydrophobic effects involved in TM protein biogenesis is sufficient for state-of-the art topology predictions of alpha-helical membrane proteins.

54. Tsirigos K, Hennerdal A, Kall L, Elofsson A: **A guideline to proteome-wide alpha-helical membrane protein topology predictions**. *Proteomics* 2012, **12**:2282-2294.

55. Peters C, Elofsson A: **Why is the biological hydrophobicity scale more accurate than earlier experimental hydrophobicity scales?** *Proteins* 2014, **82**:2190-2198.

56. Sonnhammer E, von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences**. *Proc Int Conf Intell Syst Mol Biol* 1998, **6**:175-182.

57. Tusnady G, Simon I: **The HMMTOP transmembrane topology prediction server**. *Bioinformatics* 2001, **17**:849-850.

58. Viklund 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.

59. Rost B: **PHD: predicting one-dimensional protein structure by profile-based neural networks**. *Methods Enzymol* 1996, **266**:525-539.

60. Jones D: **Improving the accuracy of transmembrane protein topology prediction using evolutionary information**. *Bioinformatics* 2007, **23**:538-544.

61. Viklund H, Elofsson A: **OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar**. *Bioinformatics* 2008, **24**: 1662-1668.

62. Bernsel A, Viklund H, Hennerdal A, Elofsson A: **TOPCONS: consensus prediction of membrane protein topology**. *Nucleic Acids Res* 2009, **37(Web Server issue)**:W465-W468.

63. • Dobson L, Remenyi I, Tusnady G: **CCTOP: a consensus constrained TOPology prediction web server**. *Nucleic Acids Res* 2015, **43**:W408-W412.
A consensus method for topology predictions.

64. •• Tsirigos K, Peters C, Shu N, Kall L, Elofsson A: **The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides**. *Nucleic Acids Res* 2015, **43**: W401-W407.
In our benchmarks this is the best method for proteome-wide alpha-helical TM proteins topology predictions.

65. Lao D, Arai M, Ikeda M, Shimizu T: **The presence of signal peptide significantly affects transmembrane topology prediction**. *Bioinformatics* 2002, **18**:1562-1566.

66. Kall L, Krogh A, Sonnhammer E: **A combined transmembrane topology and signal peptide prediction method**. *J Mol Biol* 2004, **338**:1027-1036.

67. Kall L, Krogh A, Sonnhammer E: **An HMM posterior decoder for sequence feature prediction that includes homology information**. *Bioinformatics* 2005, **21(Suppl. 1)**:i251-7.

68. Viklund H, Bernsel A, Skwark M, Elofsson A: **SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology**. *Bioinformatics* 2008, **24**:2928-2929.

69. Reynolds SM, Käll L, Riffle ME, Bilmes JA, Noble WS: **Transmembrane topology and signal peptide prediction using dynamic bayesian networks**. *PLoS Comput Biol* 2008, **4**: e1000213.

70. Zhai Y, Saier M Jr: **The beta-barrel finder (BBF) program allowing identification of outer membrane beta-barrel proteins encoded within prokaryotic genomes**. *Protein Sci* 2002, **11**:2196-2207.

71. Berven F, Flikka K, Jensen H, Eidhammer I: **BOMP: a program to predict integral beta-barrel outer membrane proteins encoded within genomes of gram-negative bacteria**. *Nucleic Acids Res* 2004, **32(Web Server issue)**:W394-9.

72. Bagos P, Liakopoulos T, Spyropoulos I, Hamodrakas S: **PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins**. *Nucleic Acids Res* 2004, **32(Web Server issue)**:W400-4.

73. Bigelow H, Petrey D, Liu J, Przybylski D, Rost B: **Predicting transmembrane beta-barrels in proteomes**. *Nucleic Acids Res* 2004, **32**:2566-2577.

74. Hayat S, Elofsson A: **BOCTOPUS: improved topology prediction of transmembrane beta barrel proteins**. *Bioinformatics* 2012, **28**:516-522.

75. • Hayat S, Peters C, Shu N, Tsirigos K, Elofsson A: **Inclusion of dyad-repeat pattern improves topology prediction of transmembrane beta-barrel proteins**. *Bioinformatics* 2016, **32**:1571-1573.
In our benchmarks this is one of the best methods for topology prediction of beta-barrel proteins.

76. • Tsirigos K, Elofsson A, Bagos P: **PRED-TMBB2: improved topology prediction and detection of beta-barrel outer membrane proteins**. *Bioinformatics* 2016, **32**:i665-i671.
In our benchmarks this is one of the best methods for topology prediction and proteome-wide discrimination of beta-barrel proteins.

77. Bagos P, Liakopoulos T, Hamodrakas S: **Evaluation of methods for predicting the topology of beta-barrel outer membrane proteins and a consensus prediction method**. *BMC Bioinform* 2005, **6**:7.

78. • Remmert M, Linke D, Lupas A, Soding J: **HHomp — prediction and classification of outer membrane proteins**. *Nucleic Acids Res* 2009, **37(Web Server issue)**:W446-51.
The best method to identify beta-barrel proteins.

79. Weigt M, White R, Szurmant H, Hoch J, Hwa T: **Identification of direct residue contacts in protein–protein interaction by message passing**. *Proc Natl Acad Sci U S A* 2009, **106**:67-72.

80. Marks D, Colwell L, Sheridan R, Hopf T, Pagnani A, Zecchina R, Sander C: **Protein 3d structure computed from evolutionary sequence variation**. *PLoS ONE* 2011, **6**:e28766.

81. Schug A, Weigt M, Onuchic J, Hwa T, Szurmant H: **High-resolution protein complexes from integrating genomic information with molecular simulation**. *Proc Natl Acad Sci U S A* 2009, **106**:22124-22129.

82. Michel M, Skwark M, Menendez Hurtado D, Ekeberg M, Elofsson A: **Predicting accurate contacts in thousands of Pfam domain families using PconsC3**. *Bioinformatics* 2017, **33**:2859-2866.

83. Wang S, Sun S, Li Z, Zhang R, Xu J: **Accurate de novo prediction of protein contact map by ultra-deep learning model**. *PLoS Comput Biol* 2017, **13**:e1005324.

84. Michel M, Menendez-Hurtado D, Uziela K, Elofsson A: **Large-scale structure prediction enabled by reliable model quality assessment and improved contact predictions for small families**. *Bioinformatics* 2017, **33**:i23-i29.

85. Ovchinnikov S, Park H, Varghese N, Huang P, Pavlopoulos G, Kim D, Kamisetty H, Kyrpides N, Baker D: **Protein structure determination using metagenome sequence data**. *Science* 2017, **355**:294-298.