REVIEW

State-of-the-art in membrane protein prediction

Chien Peter Chen[1] and Burkhard Rost[1,2]

[1]CUBIC, Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA
[2]Columbia University Center for Computational Biology and Bioinformatics, New York, NY, USA

Abstract: Membrane proteins are crucial for many biological functions and have become attractive targets for pharmacological agents. About 10%–30% of all proteins contain membrane-spanning helices. Despite recent successes, high-resolution structures for membrane proteins remain exceptional. The gap between known sequences and known structures calls for finding solutions through bioinformatics. While many methods predict membrane helices, very few predict membrane strands. The good news is that most methods for helical membrane proteins are available and are more often right than wrong. The best current prediction methods appear to correctly predict all membrane helices for about 50%–70% of all proteins, and to falsely predict membrane helices for about 10% of all globular proteins. The bad news is that developers have seriously overestimated the accuracy of their methods. In particular, while simple hydrophobicity scales identify many membrane helices, they frequently and incorrectly predict membrane helices in globular proteins. Additionally, all methods tend to confuse signal peptides with membrane helices. Nonetheless, wet-lab biologists can reach into an impressive toolbox for membrane protein predictions. However, the computational biologists will have to improve their methods considerably before they reach the levels of accuracy they claim.

Keywords: genome sequence analysis, protein structure prediction, multiple alignments, transmembrane helices, transmembrane prediction

Abbreviations:
ALOM2, hydrophobicity-based prediction of membrane helices using a discriminant function (Klein et al 1985); DAS, dense alignment surface method predicting membrane helices (Cserzö et al 1997); GES, Goldman, Engelman and Steitz (Engelman et al 1986); GPCR, G-protein coupled receptor: family of proteins with seven transmembrane helices; HMM, Hidden Markov model (statistical algorithm from machine learning); HMMTOP, Hidden Markov model predicting transmembrane helices (Tusnady and Simon 1998); KD, Kyte and Doolittle (Kyte and Doolittle 1982); KKD, application of discriminant function to the KD hydropathy (Klein et al 1985); MEMSAT, dynamic-programming-based prediction of transmembrane helices (Jones et al 1994); META-PP, internet service allowing access to a variety of bioinformatics tools through one single interface (Eyrich and Rost 2000); OM, outer membrane; PHDhtm, profile-based neural network prediction of transmembrane helices (Rost et al 1995; Rost 1996; Rost et al 1996b); PHDpsihtm, PSI-BLAST profile-based neural network prediction of transmembrane helices (Rost et al 1995; Rost 1996; Rost et al 1996b); PP (PredictProtein), internet server for protein sequence analysis and protein structure prediction (Rost et al 1994; Rost 1996; Rost 2000); PRED-TMR, propensity optimised hydropathy prediction of membrane helices (Pasquier et al 1999); PSI-BLAST, position specific iterated database search (Altschul et al 1997); SOSUI, hydrophobicity and amphiphilicity-based transmembrane helix prediction (Hirokawa et al 1998); SPLIT, transmembrane helix prediction (Juretic et al 1998); SRS, Sequence Retrieval System, ie a portal to simultaneously access most existing databases (Etzold and Argos 1993; Etzold et al 1996); TM, transmembrane; TMAP, alignment-based prediction of transmembrane helices (Persson and Argos 1996); TMFinder, multiple hydrophobicity-scale-based prediction of membrane helices (Deber et al 2001); TMH, transmembrane helix; TMHMM, Trans-Membrane prediction using Hidden Markov Models (Sonnhammer et al 1998); TMpred, membrane prediction based on statistical preferences (Hofmann and Stoffel 1993); TopPred, hydrophobicity-based membrane helix prediction (von Heijne 1992); URL, Uniform Resource Locator, ie address of a website; WW, transmembrane prediction based on the Wimley-White hydrophobicity scale (Jayasinghe et al 2001).

Correspondence: Burkhard Rost, CUBIC, Department of Biochemistry and Molecular Biophysics, Columbia University, 650 West 168th Street BB217, New York, NY 10032, USA; tel +1 212 305 3773; fax +1 212 305 7932; email rost@columbia.edu; http://cubic.bioc.columbia.edu/

Introduction

Helical membrane proteins constitute an important class of proteins. Membrane proteins are crucial for survival. They constitute key components for cell–cell signalling, mediate the transport of ions and solutes across the membrane, and are crucial for recognition of self (Stack et al 1995; Chapman et al 1998; Le Borgne and Hoflack 1998; Chen and Schnell 1999; Hettema et al 1999; Pahl 1999; Truscott and Pfanner 1999; Bauer et al 2000; Ito 2000; Soltys and Gupta 2000; Thanassi and Hutltgren 2000). Furthermore, the pharmaceutical industry preferably targets membrane-bound receptors (Heusser and Jardieu 1997; Bettler et al 1998; Moreau and Huber 1999; Saragovi and Gehring 2000; Sedlacek 2000). A prominent example for the pharmacological importance of membrane proteins is the large super-family of G protein-coupled receptors (GPCRs), which include receptors for hormone, neurotransmitter, growth factor, light and odour-related ligands (Dewji and Singer 1997; Hildebrand 1997). These receptors are of interest to the pharmaceutical industry as they present novel targets for drugs (Stadel et al 1997; Marchese 1999). In addition, more than 50% of prescription drugs act on GPCRs (Gudermann et al 1995; Attwood et al 2000; Attwood 2001). Besides the GPCRs, other important membrane protein families include ion channels, motor proteins and bio-energetically-related proteins such as those involved in the electron transport system (Kihara et al 1998; Kihara and Kanehisa 2000).

Helical membrane proteins challenge bioinformatics. Despite the great biological and medical importance of membrane proteins, we still have very little experimental information about their three-dimensional (3-D) structures. Less than 1% of the proteins of known structure are membrane proteins. High-resolution structures are scarce because membrane proteins are not easy to crystallise, and are hardly tractable by nuclear magnetic resonance (NMR) spectroscopy. Nonetheless, there are a number of recent and promising attempts to tackle membrane proteins by solid state and even by solution NMR (Fu and Cross 1999; de Groot 2000; Marassi and Opella 2000; McDermott et al 2000; Riek et al 2000; Sanders and Nagy 2000; Arora and Tamm 2001; Fernandez et al 2001; Opella et al 2001; Wuthrich 2001). Fortunately, it is relatively easy to identify the location of membrane helices through low-resolution experiments. An expert-curated list of low-resolution experiments maintained by Steffen Möller and colleagues (Möller et al 2000) considers information from C-terminal fusion with indicator proteins (McGovern et al 1991; Hennessey and Broome-Smith 1993; Traxler et al 1993; van Geest and Lolkema 2000) and from antibody-binding (Traxler et al 1993; McGuigan 1994; Jermutus et al 1998; Morris et al 1998; Amstutz et al 2001). Nevertheless, the bad news remains that we have experimental information for less than 500 helical membrane proteins. We believe that the human genome alone codes for almost 10 000 helical membrane proteins (Wallin and von Heijne 1998; Krogh et al 2001; Liu and Rost 2001). Thus, bioinformatics is challenged to help bridge the information gap between what we want and what we have.

The lipid bilayer simplifies the prediction problem. Fortunately, the task to predict aspects of structure for the membrane regions of proteins is simplified by strong environmental constraints on transmembrane proteins: the lipid bilayer of the membrane reduces the degrees of freedom to such an extent that 3-D structure formation becomes almost a 2-D problem. However, this constraint does not apply to the other class of membrane proteins, which are the porin-like proteins that form pores by β-strand barrels (von Heijne 1996; Seshadri et al 1998; Buchanan 1999). Since there is not much experimental information available on different porin-like membrane proteins, it is difficult to develop prediction methods and to estimate prediction accuracy for this class.

Here, we summarise both the state-of-the-art and to some extent the history of attempts within computational biology and bioinformatics to predict a protein's transmembrane regions. We focus on the concepts and the resulting methods that are available for everyday sequence analysis, and we discuss some of the major problems and practical aspects of these methods. The major problem in the field of membrane protein prediction is the lack of experimental high-resolution data. Consequently, estimates for prediction accuracy are perhaps overly optimistic. In this paper we suggest estimates that are as realistic as possible.

Concepts for predicting TM helix location and topology

Hydrophobicity scales provide simple criteria to predict membrane helices. Transmembrane helices (TMH) can be predicted based on the distinctive patterns of hydrophobic (transmembrane) and polar (non-membrane) regions within the sequence. These patterns are as follows: (1) TM helices are predominantly apolar and between 12 and 35 residues long (Chen et al unpub). (2) Globular regions between membrane helices are typically shorter than 60 residues (Wallin and von Heijne 1998; Liu and Rost 2001). (3) Most TMH proteins have a specific distribution of the positively

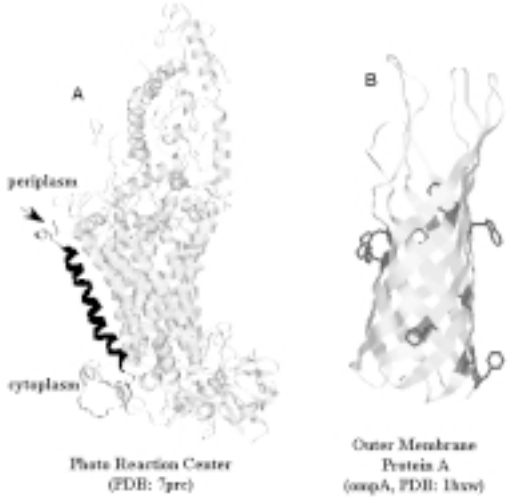

Figure 1 Two types of membrane proteins. **A:** The X-ray structure of the photo-reaction centre (PDB code 7prc) was the first high-resolution structure of an α-helical membrane protein (Deisenhofer et al 1985). Represented in light grey are α-helices and β-strands (only in the non-membrane regions). The lipid membrane bilayer is crossed by the 11 helices in the middle of the structure. The N-terminus of the H (heavy) chain is marked by an arrow (left, middle); the beginning of that chain is highlighted in dark grey (including the only membrane helix of that chain). The topology is defined by the orientation of the helices with respect to the membrane bilayer, here the upper part of the protein is located in the periplasm, the lower part in the cytoplasm. Hence, the membrane helix of the H chain has the topology OUT. **B:** The X-ray structure of the transmembrane part of the Outer Membrane Protein A (ompA, PDB: 1bxw) is an example of a β-barrel membrane protein (Pautsch and Schulz 1998). The β-strands are given in light grey, and the aromatic residues Tryptophan and Phenylalanine are in dark grey. The protein contains only half as many (8) β-strands as most porins. Typically, membrane β-strands are amphipathic, ie residues *i* and *i+2* are hydrophobic while residues *i+1* and *i+3* are hydrophilic, since one side of the strands points to the lipid bilayer and the other to the inside of the pore. We also indicated the specific band of aromatic residues that lines the interface between the core transmembrane regions and the exposed loops. The topology of beta-membrane proteins is typically determined by the location of the longest loops. Here, this loop is extracellular.

charged amino acids arginine and lysine, coined by Gunnar von Heijne as the 'positive-inside-rule' (von Heijne 1986; von Heijne 1989). Connecting 'loop' regions on the inside of the membrane have more positive charges than 'loop' regions on the outside (Figure 1). (4) Long globular regions (> 60 residues) differ in their composition from those globular regions subject to the 'inside-out-rule'. These simple facts have been at the heart of a variety of prediction methods developed over the last two decades. Methods have improved over time, and a great number of ideas have been thrown at the problem. Here, we focus on some of the major methods.

Hydrophobicity scales were introduced 20 years ago. Kyte and Doolittle (KD) developed one of the first methods that evaluated the hydrophilicity and hydrophobicity of a protein along the amino acid sequence (Kyte and Doolittle 1982). They defined a hydropathy scale that associated a hydropathy value to each amino acid. To identify membrane regions, they implemented a moving-window approach in which they simply summed the hydrophobicity scale over *w* adjacent residues in the native sequence (Kyte and Doolittle 1982). They tested window lengths of $w = 9–12$ adjacent residues, and found windows of 19 residues to discriminate best between membrane and globular proteins. Like most succeeding methods, KD then had to define some threshold *T* to label a segment as 'membrane helix': if the sum over the hydrophobicity exceeded *T*, the segment was predicted to be a membrane helix. In particular, KD suggested a threshold of $T > 1.6$ for the average over 19 residues. Around the same time, Eisenberg and colleagues developed the helical hydrophobic moment as a measure of the amphiphilicity of a helix. This hydrophobic moment differed between transmembrane and globular helices, and could thus be explored to predict transmembrane regions (Eisenberg et al 1982).

Predictions improve by processing simple hydrophobicity scales. Klein, DeLisi and colleagues combined a discriminant function (similar to the one introduced by Barrantes (1975)) with the hydrophobic analysis of KD (Klein et al 1985). In particular, they applied a quadratic discriminant function to the KD hydropathy scale and summed over a window of $w = 17$ adjacent residues. Proteins with values < 0 were classified as integral membrane proteins. Nakai and Kanehisa applied the same concept of filtering the simple scales through a quadratic discriminant function in their method ALOM2 (Nakai and Kanehisa 1992). The rationale of ALOM2 is that it first tentatively evaluates the number of putative membrane helices using a low threshold of 0.5. Then it refines the predicted number by using a more stringent threshold of –2.0. After the transmembrane regions are predicted, ALOM2 applies a modified positive-inside rule developed by Hartmann, Rapoport and Lodish (Hartmann et al 1989) to predict the protein's topology, which in the realm of membrane proteins refers to the orientation of its N-terminus with respect to the lipid bilayer. Gunnar von Heijne introduced the 'positive-inside rule' reflecting the observation that non-membrane regions inside have more positively charged residues than the regions outside (von Heijne 1986). Hartmann, Rapoport and Lodish (1989) altered this rule slightly by omitting the region flanking the first helix from the compilation. After the transmembrane regions were predicted ALOM2 used this modified positive-inside rule to predict the membrane topology.

More refined indices improve predictions. Hydropathy-based methods still appear to be effective in predicting transmembrane segments. One of the drawbacks was that such methods fail to discriminate accurately between

membrane regions and highly hydrophobic globular segments. The PRED-TMR algorithm uses a standard hydrophobicity analysis with an emphasis on the detection of potential helix ends (Pasquier et al 1999). Using propensities of amino acid residues at the termini of transmembrane helices collected by the authors, PRED-TMR compiles scores for the termini of each putative segment. Based on the two termini scores, a hydropathy score of a TMH and a length constraint, Pasquier developed a scoring function used to find the best prediction. In contrast, Jayasinghe et al attempted to improve hydropathy analysis by directly improving the hydropathy scales (Jayasinghe et al 2001). The commonly used hydrophobicity scales neglect the thermodynamic constraints α-helices impose on transmembrane stability. Hence, Jayasinghe et al derived a whole-residue hydropathy scale from the Wimley-White experiments that took into account the backbone constraints. Another new hydrophobicity scale was at the heart of the TMFinder method (Deber et al 2001). The scale (Liu-Deber scale) was based on the HPLC retention time of peptides with non-polar phase helicity. It measured the propensity of an amino acid to be in an alpha-helical state based on circular dichroism.

Amino acid preferences for membrane and non-membrane proteins can be used for prediction. Rather than using the observation that hydrophobic residues are abundant in transmembrane helices, we could conceive a more general strategy to infer from known membrane helices which amino acids have the highest preference for that state. Such a simple statistical evaluation was already the base for the first methods predicting secondary structure for globular proteins (Schulz 1988; Fasman 1989; Rost and Sander 2000). TMpred is one of the methods using such statistical preferences to predict membrane helices taken from an expert-compiled data set of membrane proteins (Hofmann and Stoffel 1993). TMpred combines several matrices for scoring. Juretic et al integrated multiple scales for amino acids for the prediction of transmembrane regions in their method SPLIT (Juretic et al 1993; Juretic et al 1998). The authors derived amino acid preferences for the 'state' membrane helix from a data set of integral membrane proteins with partially known secondary structure. They also extracted preferences for β-strand, turn and non-regular secondary structure based on sets of soluble proteins of known structure. The comparison with hydrophobicity plots suggested that the preference profiles were more accurate, exhibited higher resolution and had less noise. Shorter, unstable or movable membrane-helices were often missed by the hydrophobicity analyses in proteins with transport functions. In contrast, they were predicted by the combination of preferences. For instance, the N-terminal TM helices of voltage-gated ion channels and glutamate receptors were correctly identified by SPLIT.

Incorporating more information into methods improves prediction accuracy. A considerably more complex scheme for post-processing hydrophobicity scales was implemented in TopPred (von Heijne 1992). TopPred predicted the complete topology of membrane proteins by using hydrophobicity analysis, automatic generation of possible topologies and ranking these topologies by the positive-inside rule. First, the method introduced a particular sliding trapezoid window to detect segments of outstanding hydrophobicity using the GES-scale (Engelman et al 1986). The two bases of the trapezoid were chosen to be 11 and 21 residues long. The authors used the shape of a trapezoid to combine the favourable noise-reduction of a triangular window (Claverie and Daulmiere 1991) with a more physically relevant rectangular window, which represents the central non-polar region of the lipid bilayer. Next, TopPred explored the positive-inside rule. This rule simply states the observation that positively charged residues (Arg and Lys) are more abundant on the inside of membranes (von Heijne 1986). Generally, this fact allows for membrane protein topology prediction. However, TopPred went a step further by choosing the thresholds for considering a segment as membrane helix that yielded the optimal difference between the number of positively charged residues at the inside and at the outside. All these refinements implemented in TopPred led to a major improvement in prediction accuracy (von Heijne 1992). SOSUI combined a variety of physico-chemical parameters to detect transmembrane proteins (Hirokawa et al 1998). In particular, the following parameters are used to detect membrane helices: KD hydropathy, an amphiphilicity, relative and net charges, and protein length.

Increasing the complexity by implementing dynamic programming improves performance. In 1994, MEMSAT (Jones et al 1994) implemented statistical tables (log likelihoods) compiled from well-characterised membrane protein data and a dynamic programming algorithm, to recognise membrane topology models by expectation maximisation. Residues are classified as being one of five structural states as follows: L_i (inside loop), L_o (outside loop), H_i (inside helix end), H_m (helix middle), and H_o (outside helix end). Helix end caps are defined to span over four adjacent residues (one helical turn). Next the authors extracted the propensity of each amino acid for each of these five states from experimentally well-described membrane proteins.

Using these propensities, MEMSAT calculates a score relating a given sequence to a predicted topology and arrangement of membrane helices. The particular feature of MEMSAT is that it finds the optimal score through dynamic programming, ie an algorithm is also explored to find the optimal pairwise sequence alignment (Needleman and Wunsch 1970; Sellers 1974). Thus, MEMSAT finds the best out of a great number of possible predictions (Jones et al 1994).

Evolutionary information from protein families raises accuracy further. Until 1996, automatic methods based their predictions of membrane regions on the properties of single protein sequences. From predicting secondary structure for globular proteins, we know that using alignment information improves prediction accuracy significantly (Rost and Sander 1993; Rost and Sander 1994; Rost 2001). PHDhtm was the first method that used information from protein families for membrane predictions (Rost et al 1995; Rost et al 1996a; Rost et al 1996b). In the initial version, location and topology of membrane helices were simply predicted by a system of neural networks (Rost et al 1995). PHDhtm was then (Rost et al 1996a; Rost et al 1996b) refined by post-processing the neural network output through a dynamic programming-like algorithm, similar to the one introduced by Jones et al. The combination of various algorithms and multiple alignment information resulted in what is still one of the most accurate prediction methods today. TMAP was another early application of multiple sequence alignments to determine membrane-spanning segments (Persson and Argos 1996). It was based on propensity values determined for segments of 21 consecutive residues in transmembrane segments (P_m), and for the flanking four-residue caps (ends) of membrane helices (P_e). Residues with high P_m tended to be hydrophobic whereas those with high P_e tended to be basic and polar residues. The compositional difference in the protein segments exposed to the two surfaces of a membrane for twelve important residues was determined. Ratios were calculated for Asn, Asp, Gly, Phe, Pro, Trp, Tyr and Val (mostly found at the outside of membranes), and for Ala, Arg, Cys and Lys (mostly inside). The consensus over these twelve residues was used to predict topology. Multiple alignments improve prediction accuracy. However, for 20%–30% of all proteins there are no homologues in current databases (Liu and Rost 2001). In response to this situation, the so-called dense alignment surface (DAS) method was developed (Cserzö et al 1997). DAS is based on the RreM scoring matrix originally introduced to improve alignments for G-protein coupled receptors. It compares low-stringency dot-plots of the query protein against the background representing the universe of non-homologous membrane proteins using the RreM scoring matrix.

Grammatical rules reflect global aspects of membrane regions. The lipid bilayer constrains the structure of the membrane-passing regions of proteins in many ways. TMHMM pioneered building models of predicted membrane proteins considering a variety of such constraints in one consistent methodology (Sonnhammer et al 1998; Krogh et al 2001). A similar concept was implemented in HMMTOP (Tusnady and Simon 1998; Tusnady and Simon 2001). TMHMM and HMMTOP realise their models through hidden Markov models (HMMs). TMHMM implements a cyclic model with seven states for transmembrane-helix (TMH) core, TMH-caps on the N- and C-terminal sides, non-membrane regions on the cytoplasmic side, two non-membrane regions on the non-cytoplasmic side and a globular domain state in the middle of each non-membrane region. The two non-membrane regions on the non-cytoplasmic-side model short and long loops respectively, which correspond to two different membrane insertion mechanisms. In contrast, HMMTOP uses a hidden Markov model distinguishing the following five structural states: inside non-membrane region; inside TMH-cap; membrane helix; outside TMH-cap; and outside non-membrane region. Conceptually, this model is similar to the one used in MEMSAT (Jones et al 1994). It differed in the placement and interpretation of TMH-caps, which Tusnady et al interpret as not being in the membrane (Tusnady and Simon 1998).

Helical caps can be predicted by molecular dynamics. Molecular dynamics methods attempt to represent protein conformations, and have been used together with energy minimisation to simulate protein folding (Levitt and Warshel 1975; Hagler and Honig 1978; Levitt 1983; Karplus and Petsko 1990; Berendsen 1991; Dill 1993; van Gunsteren 1993). In practice, both the enormous complexity of the free parameters and the inaccuracy in experimentally determining the fundamental constants seriously hamper the success of such methods. However, they sometimes yield accurate predictions for short peptides such as membrane helices. Molecular dynamics simulations in an explicit lipid and water environment have been used to define the precise ends of TM helices (Forrest et al 1999; Sajot and Genest 2000). Molecular dynamics typically generates many possible models rather than unambiguously pointing to one single model. Briggs and colleagues present a new approach to selecting candidate models (Briggs et al 2001). They assume that neutral amino acid substitutions do not affect the stability

of a native structure but may destabilise the non-native structures. Applying this assumption to the α-helical transmembrane domains of two homodimers (human glycophorin A and human CD3-zeta), they in fact identify a single model by their simulation.

Concepts for predicting TM beta-sheet proteins

There is a structural variety of beta-membrane proteins. β-barrel membrane proteins are found in the outer membranes (OMs) of gram-negative bacteria and are likely in the OMs of mitochondria and chloroplasts. In prokaryotes, they mediate non-specific, passive transport of ions and small molecules, and can selectively pass molecules such as maltose and sucrose (Nikaido 1994; Schirmer et al 1995; Meyer et al 1997; Forst et al 1998; Schirmer 1998). In eukaryotic organelles, β-barrel membrane proteins have been suggested to be involved in voltage-dependent anion channels (Mannella 1998). This wide range of functions is associated with a wide range of structural variants: β-barrel membrane proteins with barrel sizes from small 8-stranded to large 22-stranded β-barrels and with different topologies (Schulz 2000). Of the β-barrel membrane proteins, porins are the best studied. Many porin barrels are trimers and contain 16 anti-parallel β-strands; maltoporin from *Escherichia coli* contains 18 strands (Meyer et al 1997). A band of hydrophobic residues encircles the trimer (Weiss and Schulz 1992; Pebay-Peyroula et al 1995; Meyer et al 1997) (Figure 1). Porins also contain a central channel that is partially blocked by a loop that folds inwardly and is attached to the inner side of the barrel wall (Schirmer 1998). This arrangement forms an 'eyelet', which defines the size of solute molecule that can traverse the channel. Currently, high-resolution structures are only available for bacterial OM proteins (Tamm et al 2001).

Membrane strands are difficult to predict. Unlike α-helical membrane proteins, there are no simple low-resolution experiments that yield large amounts of data for β-barrel membrane proteins. This has constrained the ability to develop prediction methods. Many β-strands contain alternating hydrophobic and hydrophilic side-chains. However, this simple rule usually does not suffice to identify membrane strands (Schulz 2000). Methods that implement physico-chemical properties were applied successfully only in the context of experimental information (Paul and Rosenbusch 1985; Welte et al 1991; Schirmer and Cowan 1993). All early attempts to predict membrane strands employed the amphipacity and hydrophobicity of β-strands. Paul and Rosenbusch attempted a minimal approach to predict and identify segments causing polypeptides to reverse their direction (turn identification), but they avoided hydrophobicity parameters (Paul and Rosenbusch 1985). In contrast, Jahnig suggested that a generalisation of hydrophobicity analysis was sufficient to predict membrane-spanning amphiphilic α-helices and β-strands (Jahnig 1990). Unfortunately, membrane strands have no long stretch of consecutive hydrophobic residues. In fact, the overall hydrophobicity for β-barrel membrane proteins is similar to that of soluble proteins. Welte and colleagues compared the hydrophilicity profiles and sequences of porin from *Rhodobacter capsulatus* with those of OmpF and PhoE from *Escherichia coli*. They determined a set of specific insertions and deletions in the alignments of these proteins, and inferred that OmpF and PhoE have similar structures in their membrane-spanning regions. Their experimental work verified this prediction (Welte et al 1991). Cowan and colleagues (Cowan et al 1992) suggested to use the mean hydrophobicity of one side of a putative β-strand by averaging over hydrophobic moments (Eisenberg et al 1984) of every second residue within a sliding window (Vogel and Jahnig 1986; Schirmer and Cowan 1993). To improve the signal-to-noise-ratio, they accounted for the band of aromatic residues in flanking positions of the β-strands. Another method that was considered for predicting β-membrane spanning regions was a rule-based approach. Gromiha and colleagues combined amino acid preferences for β-strands with the surrounding hydrophobicity of the respective residues to predict β-strands (Gromiha and Ponnuswamy 1993; Gromiha et al 1997). With their method they reproduced about 82% of the residues in structurally known membrane regions.

Non-linear statistics enables prediction of membrane β-strands. Diederichs and colleagues proposed to use a neural network to predict the topology of the bacterial OM β-strand proteins and to locate residues along the axes of the pores (Diederichs et al 1998). The neural network predicts the z-coordinate of C-alpha atoms in a coordinate frame with the outer membrane in the xy-plane, such that low z-values indicate periplasmic turns, medium z-values indicate transmembrane β-strands and high z-values indicate extracellular loops. Most recently, Jacoboni, Fariselli, Casadio and colleagues applied a method combining neural networks and dynamic programming to predict the location of membrane strands (Jacoboni et al 2001). The networks used alignment information as input, and predicted whether or not a particular residue is part of a membrane strand. In

the second step, the method simply finds the optimal path through the network prediction, much like the methods applied to predict membrane helical proteins (Jones et al 1994; Rost et al 1995; Rost et al 1996a). Finally, the topology is assigned based on the location of the longest loop that is taken to be exterior. The authors estimated that their system correctly predicts about 93% of all known membrane-strands. It is not clear whether or not the estimates from Diederichs et al and Jacoboni et al will hold true for all β-strand membrane proteins. The first problem is a merely technical one: for such a small set of experimentally known families (about 15 different families) (Bigelow and Rost, unpublished data) it is almost impossible to avoid over-training methods with many free parameters (such as neural networks). The second problem is of principle nature: we have to assume that the 15 different β-strand membrane families, for which we have high-resolution structures, are representative of all β-strand membrane proteins. This may turn out to be an incorrect assumption.

Practical Aspects

Availability

Most methods described are available through public servers. A list of URLs and the contact addresses are given in Table 1. Most programs – except for ALOM2, Eisenberg, KD, KKD, PRED-TMR, TMAP, TMpred, and WW – are also available through META-PP, which provides a single interface to simultaneously access many high-quality servers (Eyrich and Rost 2000). This concept of accessing many servers through one has been pioneered by the BCM-Launcher (Smith et al 1996), supposedly accessing the largest number of different methods. Other combinations are given by NPSA (NPSA 2001), META-Poland (Rychlewski 2000), and ProSAL (Kleywegt 2001). In contrast to all others, META-PP attempts to (1) return as few results as possible by filtering out technical messages and (2) combine only high-quality methods. A generalisation of the 'common interface' idea is implemented in the sequence retrieval system SRS (Etzold and Argos 1993; Etzold et al 1996), which enables simultaneous access of most existing databases. Successively SRS starts to also incorporate the direct access to prediction methods.

Prediction accuracy

Performance of prediction methods has been overestimated significantly! For all the methods described in this review high levels of prediction accuracy have been reported. Frequently, authors were daring enough to claim that their methods correctly predicted more than 90% of all membrane helices. We cannot estimate the accuracy of existing methods since they have all been developed using the known membrane proteins. However, we can estimate an upper limit for prediction accuracy. This limit suggests that developers have overrated their methods by 15%–50% (Chen et al unpub). How could this have happened? There are a variety of reasons. (1) We do not have enough high-resolution structures to allow a statistically significant analysis (Chen et al unpub). With this bottleneck, training/developing and

Table 1 Availability of prediction methods

Method	*Server*	*Program*
Helical membrane proteins		
ALOM	psort.nibb.ac.jp/form.html	Kenta Nakai: knakai@ims.u-tokyo.ac.jp
DAS	www.sbc.su.se/~miklos/DAS	miklos@bip.bham.ac.uk
HMMTOP	www.enzim.hu/hmmtop	Gábor E Tusnády: tusi@enzim.hu
MEMSAT	www.psipred.net	David Jones: d.jones@cs.ucl.ac.uk
KD	fasta.bioch.virginia.edu/fasta/grease.htm	William Pearson: wrp@virginia.edu
PHDhtm	cubic.bioc.columbia.edu/predictprotein	Burkhard Rost: rost@columbia.edu
SOSUI	sosui.proteome.bio.tuat.ac.jp/	Mitaku Group: sosui@proteome.bio.tuat.ac.jp
SPLIT	www.mbb.ki.se/tmap/index.html	Davor Juretic: juretic@mapmf.pmfst.hr
TMAP	www.mbb.ki.se/tmap/index.html	Bengt Persson: Bengt.Persson@ibp.vxu.se
TMHMM	www.cbs.dtu.dk/services/TMHMM-2.0	Anders Krogh: krogh@cbs.dtu.dk
TMpred	www.ch.embnet.org/software/	
TopPred2	http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html	Gunnar von Heijne: gunnar@dbb.su.se
WW	blanco.biomol.uci.edu/mpex/	Stephen White: blanco@helium.biomol.uci.edu
β-sheet membrane proteins		
β-strand predictor		www.biocomp.unibo.it (upon request)

test sets may share or have homologous members. To get around this problem, developers include low-resolution experimental data and structures in their data sets. One caveat of this practice is to assume that low-resolution experiments, eg gene fusion, are sufficiently similar to high-resolution structures (crystallography). Unfortunately, this is not the case. In fact, low-resolution experiments differ from high-resolution experiments almost as much as prediction methods do (Chen et al unpub). Hence, low-resolution experiments are not sufficient to evaluate prediction accuracy. (2) All methods optimise some parameters. Since there are so few high-resolution structures, all methods use as many of the known ones as possible. However, methods perform much better on proteins for which they were developed than on new proteins, and this was overlooked in a recent analysis of prediction methods (Möller et al 2001). (3) Methods using evolutionary information failed due to the surprising fact that membrane helices are not entirely conserved across species. This observation is surprising since it implies that these proteins either do not perform similar cellular functions, eg G-coupled receptor, or that we can actually realise the function with a different number of membrane regions in some cases. (4) Finally, levels of prediction accuracy published between methods can often not be compared appropriately to one another since they are frequently based on different measures for prediction accuracy and on different data sets. The latter prompted Möller, Apweiler and colleagues to collect a set of well-characterised integral membrane proteins (Möller et al 2000). Each protein has been assigned a reliability index depending on the available structural and biochemical data. Currently, from the total set of 320 proteins in the data set, there are 33 membrane proteins with known structures, 24 with biochemical characterisation and 142 with partial biochemical evidence. The data set can be accessed via ftp://ftp.ebi.ac.uk/databases/testsets/transmembrane.

Most methods get the number of helices right for most membrane proteins. All methods based on advanced algorithms tend to underestimate transmembrane helices (Table 2: *%obs > %prd*). Thus, about 86% of the TMH residues predicted by the best methods in this category (PHDhtm and DAS) are correctly predicted. Assume that we consider a prediction of a membrane helix correct if the predicted and the observed helical regions differ by less than three residues. Given this measure for accuracy, we found that the best current methods correctly predict all membrane

Table 2 Accuracy of popular prediction methods[a]

	Per-segment accuracy[c]				*Per-residue accuracy*[d]				
Method[b]	Q_{ok}	$Q_{htm}^{\%obs}$	$Q_{htm}^{\%prd}$	*TOPO*	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%prd}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%prd}$
ERROR[e]	± 9	± 7	± 7	± 9	± 3	± 6	± 7	± 4	± 4
DAS	79	**99**	96		*72*	*48*	**94**	**97**	*62*
HMMTOP2	83	**99**	**99**	61	79	70	89	*88*	*71*
PHDhtm07	86	**99**	**98**	50	**80**	72	87	*82*	74
PHDhtm08	86	**99**	**98**	54	**80**	72	87	*82*	74
PHDpsihtm07	84	**99**	**98**	66	**80**	**76**	83	*86*	**80**
PRED-TMR	*61*[f]	*84*	90		76	*58*	85	94	*66*
SOSUI	71	88	86		75	*66*	*74*	*80*	*69*
TMHMM1	71	90	90	45	**80**	*68*	81	89	*72*
TopPred2	75	90	90	54	77	*64*	83	90	*69*
WW	*54*	95	91		*71*	71	*72*	*67*	*67*

[a] Data set: Sequence-unique subset of 36 high-resolution membrane helical proteins from PDB (Berman et al 2000). Note: this is the largest subset of all 105 high-resolution membrane chains, which fulfils the condition that no pair in the set has significant sequence similarity as defined in Rost (1999).

[b] Methods: see abbreviations at begin of article.

[c] Per-segment accuracy: Q_{ok} percentage of proteins for which all TM helices are predicted correctly (allowed deviation of up to 3 residues), $Q_{htm}^{\%obs}$ percentage of all observed helices that are correctly predicted, $Q_{htm}^{\%prd}$ percentage of all predicted helices that are correctly predicted, *TOPO* percentage of proteins for which the topology (orientation of helices) is correctly predicted (note: empty for methods that do not predict topology).

[d] Per-residue accuracy: Q_2 percentage of correctly predicted residues in two-states: membrane helix / non-membrane helix, $Q_{2T}^{\%obs}$ percentage of all observed TMH helix residues that are correctly predicted, $Q_{2T}^{\%prd}$ percentage of all predicted TMH helix residues that are correctly predicted, $Q_{2N}^{\%obs}$ percentage of all observed non-TMH helix residues that are correctly predicted, $Q_{2N}^{\%prd}$ percentage of all predicted non-TMH helix residues that are correctly predicted.

[e] ERROR: the estimates for per-segment accuracy resulted from a bootstrap experiment with M = 100 and K = 18; the estimates for per-residue accuracy were obtained by standard deviations over Gaussian distributions for the respective score.

[f] Numbers in italics: two standard deviations below the numerically highest value in each column (set in bold letters).

NOTE: all methods are tested on the same set of proteins. However, the numbers are NOT from a cross-validation experiment, ie some methods may have used some of the proteins for training. Generally, newer methods are more likely to be overestimated than older ones. In particular, HMMTOP2, TMHMM1, and WW have been developed using ALL the proteins, listed here.

helices correctly for 70%–75% of all proteins (Table 2). However, the topology is predicted correctly for only about half of all proteins. The only exception is HMMTOP2, but all proteins tested here were used to train HMMTOP2, hence the level of 61% accuracy in topology prediction may be overestimated significantly. In terms of per-residue scores, the best current methods correctly predict more than 65% of the observed TMH residues correctly (Table 2). Although the results summarised in Table 2 are similar to those recently compiled on a non-unique set of low- and high-resolution structures (Möller et al 2001), most estimates still constitute overestimates since very few methods shown (DAS, PHDhtm, TopPred2) did NOT use most of the proteins to optimise prediction accuracy.

Simple hydrophobicity scales are less accurate than advanced methods. A surprising result recently published suggested that simple hydrophobicity scales predict membrane helices almost as accurately as do the most advanced current prediction methods (Möller et al 2001). We tested 20 different hydrophobicity scales on various data sets and could not confirm this optimism (Chen et al unpub). Rather, the example given in Table 2 for the Whitney-White scale (WW) appeared to be one of the best simple hydrophobicity scales, although it predicts all membrane helices correctly for only 54% of the proteins tested. In fact, most hydrophobicity scores locate all helices without over-prediction for less than 40% of the proteins (Chen et al unpub).

All methods confuse membrane helices with signal peptides. Signal peptides that are cleaved off secreted proteins usually contain stretches of hydrophobic residues resembling membrane helices (Nielsen et al 1996; Nielsen et al 1997a; Nielsen et al 1997b; Nielsen et al 1999). Hence, most methods confuse signal peptides with membrane helices. The best separation is achieved by ALOM2, a method optimised to sort proteins into classes of sub-cellular localisation (Nakai and Kanehisa 1992; Nakai and Horton 1999). The most accurate specialists for membrane prediction (TMHMM and PHDhtm) appear to falsely predict signal peptides as membrane helices for 30%–40% of all the signal peptides we tested (Chen et al unpub). Surprisingly accurate in rejecting signal peptides is the Wolfenden scale for hydrophobicity (Wolfenden et al 1979). All other hydrophobicity scales predict more than 90% of the signal peptides as membrane helices (Chen et al unpub).

Many methods predict membrane helices in globular proteins. Interestingly, most methods have also been overestimated significantly in their ability to distinguish between globular and membrane proteins. Particularly poor is the distinction by hydrophobicity-based methods, which have reached levels of nearly 100% false positives (Chen et al unpub). In fact, the only scales we tested that incorrectly detected membrane helices in less than 80% of all globular proteins we tested (Chen et al unpub) were: Wolfenden = 2%, WW = 32%, and Eisenberg-scale = 66%. SOSUI, TMHMM1 and PHDhtm currently distinguish best between membrane and non-membrane proteins. These three predict membrane helices in less than 2% of the globular proteins. Similar results were reported on globular proteins taken from SWISS-PROT (Möller et al 2001).

Genome analysis

Despite the overestimated performance, predictions of transmembrane helices are valuable tools to quickly scan proteomes of entirely sequenced organisms for membrane proteins. As stated above, hydrophobicity-based methods mostly fail to distinguish membrane and globular proteins (Chen et al unpub). Nevertheless, the averages of helical membrane proteins published for entire genomes are surprisingly similar between different authors (Goffeau et al 1993; Rost et al 1996b; Arkin et al 1997; Frishman and Mewes 1997; Jones 1998; Wallin and von Heijne 1998; Liu and Rost 2001). Apparently, about 10%–30% of all proteins contain membrane helices. One crucial difference between the results from different groups is that more cautious estimates do not find a statistically significant difference in the percentages of TMH proteins between the three kingdoms: eukaryotes, prokaryotes and archae (Liu et al unpub). Thus, the overall content of helical membrane proteins appears not to correlate with the postulated complexity of an organism (eukaryotes more complex than prokaryotes; prokaryotes more complex than archae). However, eukaryotes have significantly more proteins with over 10 membrane helices than all other species. Furthermore, the three kingdoms also differ in the types of membrane proteins that are most abundant. For example, eukaryotes have more 7TM proteins (receptors), while prokaryotes have more 6- and 12TM proteins (ABC transporters) (Wallin and von Heijne 1998; Liu and Rost 2001).

Emerging and future developments

Membrane-helix predictions can be improved by averaging over many methods. The prediction of secondary structure for globular proteins can be improved by combining many prediction methods (Rost 2001; Rost et al 2001). Applying a

similar average, Promponas and colleagues developed their method CoPreTHi, a web-based application that uses the results from DAS, ISREC-SAPS, PHDhtm, PRED-TMR, SOSUI, TMpred and TopPred2 (Promponas et al 1999). CoPreTHi combines the results into a joint prediction histogram; residues are predicted as transmembrane if they are identified as such by at least three methods. Nilsson and colleagues explored consensus predictions for membrane protein topology to derive a reliability for the prediction (Nilsson et al 2000). In particular, they used five methods (TMHMM, HMMTOP, MEMSAT, TopPred2, and PHDhtm) to evaluate a test set of 60 *Escherichia coli* inner membrane proteins with experimentally determined topologies. They found that prediction performance varies strongly with the number of methods that agree, and that the topology of nearly half of all inner membrane proteins can be predicted with high reliability (>90% correct predictions) by a simple majority vote. When only two methods agree on topology, none of the topologies were found to be correct.

Identifying amphiphilic α-helices may improve predictions. A number of α-helix forming peptides have been reported to promote membrane fusion and other biological events related to the disruption of the hydrophobic/hydrophilic interface induced by the hydrophobicity gradient along the central helical axis. This hydrophobicity gradient may facilitate the penetration of a membrane, and may thus destabilise the packing of the lipids in the membrane bilayer and/or of the protein/water interface. This could then disrupt the interface and promote related biological events (Martin et al 1994; Brasseur et al 1997; Fujii 1999; Pecheur et al 1999; Peuvot et al 1999; Brasseur 2000). To facilitate more detailed descriptions of amphiphilic α-helices, quantitative methods have been developed that measure the overall amphiphilicity of helices. Examples are the Depth Weighted Insertion Hydrophobicity (DWIH) method (Roberts et al 1997) and the commonly used hydrophobic moment introduced by Eisenberg and colleagues (Eisenberg 1982; Eisenberg et al 1984). Harris et al (2000) improved the identification of obliquely orientated α-helices through a hydrophobic moment plot. In particular, they found a linear association between the mean hydrophobic moment $<\mu_H>$ and the corresponding mean hydrophobicity, $<H_0>$. The association was described by the least squares regression line: $<\mu_H> = 0.508 - 0.422 < H_0 >$. Hence, proteins that fall along this line would be a putative oblique-orientated α-helix. The results suggested that oblique orientated α-helices may possess a characteristic balance between the amphiphilicity and the hydrophobicity of their structures (Harris et al 2000).

Helical-membrane and signal peptide predictions have to be combined explicitly. One of the problems with some of the current methods is that they falsely predict signal peptides as transmembrane helices. The best signal peptide identification tool appears to be SignalP (Nielsen et al 1997a; Nielsen et al 1997b; Nielsen et al 1999). Trivially, this method can be incorporated into a post-prediction filter to remove predicted helices in the signal peptide region. Two methods have been developed that work in this direction. (1) PSORT (Nakai and Horton 1999) uses a variety of predictions and sequence motifs to group proteins according to their sub-cellular localisation, thereby implicitly combining membrane predictions and signal peptide predictions. (2) HMMTOP and TMHMM implicitly use known signal peptides to refine their predictions. However, a more thorough combination is still missing.

There are databases for particular families of membrane proteins and sequence motifs. Databases of protein signatures, ie relatively short sequence motifs, are becoming increasingly valuable diagnostic resources. While PROSITE (Hofmann et al 1999) annotates single motifs that have been unravelled experimentally, PRINTS encodes groups of motifs in the form of fingerprints (Attwood et al 2000). For instance, receptor subtype fingerprints comprise different parts of the terminal, loop and TM regions of G-protein-coupled receptors (GPCRs). Databases such as these can certainly be incorporated into membrane protein prediction methods to help identify novel receptors (Horn et al 2001). The strong interest in GPCRs has also led to specialised bioinformatics tools that identify GPCRs. Kim et al (2000) presented an algorithm dubbed quasi-periodic feature classifier (QFC) that characterises the physico-chemical properties of membrane proteins with multiple helices. They apply a non-parametric linear discriminant function to their variables describing the 'feature space', and thus separate GPCRs from non-GPCRs. The expected advantage of this approach is that it may find more remotely similar homologues than methods purely based on sequence similarity. Unfortunately, a thorough cross-validation of the method that would undermine this hope is still missing.

Membrane-specific substitution matrices improve database searches. Database searches are based on alignment methods that need to score the match of amino acid X in protein A with amino acid Y in protein B. A variety of substitution matrices are used for this purpose (Barton 1996; Henikoff 1996). All these substitution matrices were developed based on data sets of globular proteins. It is then not surprising that these matrices are not optimal to align

membrane regions. Ng, Henikoff and Henikoff (2000) have recently addressed this problem by developing the membrane-helix specific substitution matrix PHAT. They demonstrated that this matrix aligned membrane proteins more accurately than globular matrices. The PHAT matrix series used target frequencies from PHDhtm matrices (ie from transmembrane regions) and background frequencies from the Persson-Argos matrix (ie from hydrophobic regions) with corresponding relative entropy. Obviously, the necessary next step is to implement the following cycle: (1) predict membrane helices based on standard alignments, (2) use PHAT for the predicted membrane region to realign, (3) use the PHAT alignment to refine the prediction, and (4) possibly repeat steps 2–3. Such a refined search may allow automatic detection of distant similarities in the twilight zone (Rost 1999) that otherwise remain hidden until the experimental structure is available.

Conclusions

Optimist: membrane predictions are relatively accurate and useful. Overall, prediction methods are more accurate and more useful for membrane proteins than they are for globular proteins. The best current methods for helical membrane proteins appear to correctly predict all membrane helices for more than 60% of all proteins. Furthermore, all advanced methods that are not based solely on hydrophobicity incorrectly detect membrane helices in less than 10% of all globular proteins (Möller et al 2001; Chen et al unpub). In contrast, most methods based only on hydrophobicity go wrong for more than 80% of all globular proteins, and even the best current methods frequently confuse signal peptides and membrane helices (Chen et al unpub). Nevertheless, most often the best methods correctly reject signal peptides (Chen et al unpub). Most prediction errors constitute the over- or under-prediction of a single membrane helix. While this has important impacts on functionally classifying the protein, the good news is again that most often the good methods correctly predict the number of membrane helices, ie they may help in providing a first clue about aspects of function in the context of genome analysis. Recently, a number of tools have addressed the problem to predict β-membrane proteins. The estimated levels of prediction accuracy are promising. Unfortunately, there is no accurate method yet that detects β-membrane proteins in context of entire genome searches.

Pessimist: all methods have been overestimated significantly. A number of recently determined high-resolution structures of membrane proteins revealed that the accuracy of low-resolution experiments may have been overestimated. However, the accuracy of prediction methods was overestimated more seriously. Particular problems for prediction methods result from the following observations (Chen et al unpub): (1) many membrane helices span over more than 30 residues, and (2) membrane helices are not as well-conserved as they appeared to be in the much smaller sequence databases of a decade ago. A seemingly 'trivial' flaw of many estimates published by various groups was that they compared results based on different data sets and different scores measuring accuracy. The most important problem may have been that developers were not careful enough in avoiding overfitting the few experimentally known proteins. In fact, this reality strongly constrains the estimates of accuracy provided in this analysis: only the methods published before 1998 (DAS, PHD, and TopPred2) did not use most of the proteins for which the results are given in Table 2. Thus, the actual prediction accuracy may even be lower. Current prediction methods are still valuable both for everyday sequence and entire proteome analysis. However, it seems that 'simple' predictions of the location of helices are not as simple as anticipated. A lot of work remains to be done before we reach the levels of accuracy that optimists may have believed were reached a decade ago.

The ultimate solution: we need more high-resolution experiments! Promising new strategies may yield more high-resolution structures of membrane proteins. The frequently observed instability of membrane proteins outside of a lipid-bilayer may call for crystallising these proteins in membrane-like environments (Landau and Rosenbusch 1996). Such a membrane system, which consists of lipid, water and protein in appropriate proportions, forms a complex three-dimensional lipidic array providing nucleation sites (seeding), and supports growth by lateral diffusion of protein molecules in the membrane (feeding). Future developments may include the use of different lipids, the inclusion of various additives, the development of different types of crystallisation screens and the rational introduction of covalent or non-covalent lattice contacts (Gouaux 1998). Although, structural genomics for membrane proteins is still far away, we hope that with every dozen new high-resolution structures solved prediction methods will gradually evolve. How many years will it then take until prediction methods reach the levels of accuracy that have been mistakenly published already in the last millennium? The answer depends on the number of surprises about non-canonical features of membrane proteins that await us on the road ahead! Clearly, the surprise details in high-resolution structures over the last five years have re-opened the field of simply predicting the topology and location of membrane proteins.

Acknowledgements

Thanks to Jinfeng Liu (Columbia) for computer assistance and to Henry Bigelow for comments and help with β-membrane prediction methods. Particular thanks to Volker Eyrich (Columbia) for programming and maintaining most of the immensely valuable software that runs the EVA and META-PredictProtein servers! The work of Burkhard Rost was supported by the grants 1-P50-GM62413-01 and RO1-GM63029-01 from the National Institute of Health. Finally, thanks to all those who enable the development of prediction methods by depositing experimental information in public databases, and to all those who maintain such databases.

References

Altschul S, Madden T, Shaffer A, Zhang J, Zhang Z, Miller W, Lipman D. 1997. Gapped Blast and PSI-Blast: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389–402.

Amstutz P, Forrer P, Zahnd C, Pluckthun A. 2001. In vitro display technologies: novel developments and applications. *Curr Opin Biotechnol*, 12:400–5.

Arkin IT, Brünger AT, Engelman DM. 1997. Are there dominant membrane protein families with a given number of helices? *Proteins*, 28:465–6.

Arora A, Tamm LK. 2001. Biophysical approaches to membrane protein structure determination. *Curr Opin Struct Biol*, 11:540–7.

Attwood TK. 2001. A compendium of specific motifs for diagnosing GPCR subtypes. *Trends Pharmacol Sci*, 22:162–5.

Attwood TK, Croning MD, Flower DR, Lewis AP, Mabey JE, Scordis P, Selley JN, Wright W. 2000. PRINTS-S: the database formerly known as PRINTS. *Nucleic Acids Res*, 28:225–7.

Barrantes F. 1975. The nicotinic cholinergic receptor: different compositions evidenced by statistical analysis. *Biochem Biophys Res Commun*, 62:407–14.

Barton GJ. 1996. Protein sequence alignment and database scanning. In Sternberg MJE, ed. Protein structure prediction. Oxford: Oxford Univ Pr. p 31–64.

Bauer MF, Hofmann S, Neupert W, Brunner M. 2000. Protein translocation into mitochondria: the role of TIM complexes. *TICB*, 10:25–31.

Berendsen HJC. 1991. Molecular dynamics studies of proteins and nucleic acids. *Curr Opin Struct Biol*, 1:191–5.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res*, 28:235–42.

Bettler B, Kaupmann K, Bowery N. 1998. GABAB receptors: drugs meet clones. *Curr Opin Neurobiol*, 8:345–50.

Brasseur R. 2000. Tilted peptides: a motif for membrane destabilization (hypothesis). *Mol Membr Biol*, 17:31–40.

Brasseur R, Pillot T, Lins L, Vandekerckhove J, Rosseneu M. 1997. Peptides in membranes: tipping the balance of stability. *TIBS*, 22:167–71.

Briggs JA, Torres J, Arkin IT. 2001. A new method to model membrane protein structure based on silent amino acid substitutions. *Proteins*, 44:370–5.

Buchanan SK. 1999. β-Barrel proteins from bacterial outer membranes: Structure, function and refolding. *Curr Opin Struct Biol*, 9:455–61.

Chapman R, Sidrauski C, Walter P. 1998. Intracellular signaling from the endoplasmic reticulum to the nucleus. *Annu Rev Cell Dev Biol*, 14:459–85.

Chen X, Schnell DJ. 1999. Protein import into chloroplasts. *Trends Cell Biol*, 9:222–7.

Chen CP, Kernytsky A, Rost B. Myths of transmembrane helix predictions. In possession of the author.

Claverie J-M, Daulmiere C. 1991. Smoothing profiles with sliding windows: better to wear a hat! *CABIOS*, 7:113–15.

Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. 1992. Crystal structure explain functional properties of two E. coli porins. *Nature*, 358:727–33.

Cserzö M, Wallin E, Simon I, von Heijne G, Elofsson A. 1997. Prediction of transmembrane α-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engin*, 10:673–6.

de Groot HJ. 2000. Solid-state NMR spectroscopy applied to membrane proteins. *Curr Opin Struct Biol*, 10:593–600.

Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, Muskat BL, Cuticchia AJ. 2001. TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci*, 10:212–19.

Deisenhofer J, Epp O, Mii K, Huber R, Michel H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution. *Nature*, 318:618–24.

Dewji NN, Singer SJ. 1997. The seven transmembrane spanning topography of the Alzheimer disease-related presenilin proteins in the plasma membranes of cultured cells. *Proc Natl Acad Sci USA*, 94:14024–30.

Diederichs K, Freigang J, Umhau S, Zeth K, Breed J. 1998. Prediction by a neural network of outer membrane beta-strand protein topology. *Protein Sci*, 7:2413–20.

Dill KA. 1993. Folding proteins: finding a needle in a haystack. *Curr Opin Struc Biol*, 3:99–103.

Eisenberg D, Schwartz E, Komaromy M, Wall R. 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J Mol Biol*, 179:125–42.

Eisenberg D, Weiss RM, Terwilliger TC. 1982. The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, 299:371–4.

Engelman DM, Steitz TA, Goldman A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Biophys Chem*, 15:321–53.

Etzold T, Argos P. 1993. SRS – an indexing and retrieval tool for flat file data libraries. *Comput Appl Biosci*, 9:49–57.

Etzold T, Ulyanov A, Argos P. 1996. SRS: Information retrieval system for molecular biology data banks. *Meth Enzymol*, 266:114–28.

Eyrich V, Rost B. 2000. The META-PredictProtein server [online]. CUBIC, Columbia University, Department of Biochemistry and Molecular Biophysics. Accessed 12 February 2002. URL: http://cubic.bioc.columbia.edu/predictprotein/submit_meta.html

Fasman GD. 1989. The development of the prediction of protein structure. In Fasman GD, ed. Prediction of protein structure and the principles of protein conformation. New York, London: Plenum Pr. p 193–303.

Fernandez C, Hilty C, Bonjour S, Adeishvili K, Pervushin K, Wuthrich K. 2001. Solution NMR studies of the integral membrane proteins OmpX and OmpA from *Escherichia coli*. *FEBS Lett*, 504:173–8.

Forrest LR, Tieleman DP, Sansom MS. 1999. Defining the transmembrane helix of M2 protein from influenza A by molecular dynamics simulations in a lipid bilayer. *Biophys J*, 76:1886–96.

Forst D, Welte W, Wacker T, Diederichs K. 1998. Structure of the sucrose-specific porin ScrY from *Salmonella Typhimurium* and its complex with sucrose. *Nat Struct Biol*, 5:37–46.

Frishman D, Mewes HW. 1997. Protein structural classes in five complete genomes. *Nat Struct Biol*, 4:626–8.

Fu R, Cross TA. 1999. Solid-state nuclear magnetic resonance investigation of protein and polypeptide structure. *Annu Rev Biophys Biophys Chem*, 28:235–68.

Fujii G. 1999. To fuse or not to fuse: the effects of electrostatic interactions, hydrophobic forces and structural amphiphilicity on protein-mediated membrane destabilization. *Adv Drug Deliv Rev*, 38:257–77.

Goffeau A, Slonimski P, Nakai K, Risler JL. 1993. How many yeast genes code for membrane-spanning proteins? *Yeast*, 9:691–702.

Gouaux E. 1998. It's not just a phase: crystallization and x-ray structure determination of bacteriorhodopsin in lipidic cubic phases. *Structure*, 15:5–10.

Gromiha MM, Majumdar R, Ponnuswamy PK. 1997. Identification of membrane spanning beta strands in bacterial porins. *Protein Engin*, 10:497–500.

Gromiha MM, Ponnuswamy PK. 1993. Prediction of transmembrane beta-strands from hydrophobic characteristics of proteins. *Int J Pept Protein Res*, 42:420–31.

Gudermann T, Nurnberg, B, Schultz G. 1995. Receptors and G proteins as primary components of transmembrane signal transduction. Part 1. G-protein-coupled receptors: structure and function. *J Mol Med*, 73:51–63.

Hagler AT, Honig B. 1978. On the formation of protein tertiary structure on a computer. *Proc Natl Acad Sci USA*, 75:554–8.

Harris F, Wallace J, Phoenix DA. 2000. Use of hydrophobic moment plot methodology to aid the identification of oblique orientated alpha-helices. *Mol Membr Biol*, 17:201–7.

Hartmann E, Rapoport TA, Lodish, HF. 1989. Predicting the orientation of eukaryotic membrane spanning proteins. *Proc Natl Acad Sci USA*, 86:5786–90.

Henikoff S. 1996. Scores for sequence searches and alignments. *Curr Opin Struct Biol*, 6:353–60.

Hennessey ES, Broome-Smith JK. 1993. Gene-fusion techniques for determining membrane-protein topology. *Curr Opin Struct Biol*, 3:524–31.

Hettema EH, Distel B, Tabak HF. 1999. Import of proteins into peroxisomes. *Biochim Biophys Acta*, 1451:17–34.

Heusser C, Jardieu P. 1997. Therapeutic potential of anti-IgE antibodies. *Curr Opin Immunol*, 9:805–13.

Hildebrand JGS, Shepherd GM. 1997. Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annu Rev Neurosci*, 20:595–631.

Hirokawa T, Boon-Chieng S, Mitaku S. 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14:378–9.

Hofmann K, Bucher P, Falquet L, Bairoch A. 1999. The PROSITE database, its status in 1999. *Nucleic Acids Res*, 27:215–19.

Hofmann K, Stoffel W. 1993. TMBASE – a database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler*, 374:166.

Horn F, Vriend G, Cohen FE. 2001. Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. *Nucleic Acids Res*, 29:346–9.

Ito A. 2000. Mitochondrial processing peptidase: multiple-site recognition of precursor proteins. *TICB*, 10:25–31.

Jacoboni I, Martelli PL, Fariselli P, De Pinto V, Casadio R. 2001. Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural network-based predictor. *Protein Sci*, 10:779–87.

Jahnig F. 1990. Structure predictions of membrane proteins are not that bad. *TIBS*, 15:93–5.

Jayasinghe S, Hristova K, White SH. 2001. Energetics, stability, and prediction of transmembrane helices. *J Mol Biol*, 312:927–34.

Jermutus L, Ryabova LA, Pluckthun A. 1998. Recent advances in producing and selecting functional proteins by using cell-free translation. *Curr Opin Biotechnol*, 9:534–48.

Jones DT. 1998. Do transmembrane protein superfolds exist? *FEBS Lett*, 423:281–5.

Jones DT, Taylor WR, Thornton JM. 1994. A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochem*, 33:3038–49.

Juretic D, Zucic D, Lucic B, Trinajstic N. 1998. Preference functions for prediction of membrane-buried helices in integral membrane proteins. *Comput Chem*, 22:279–94.

Juretic D, Lee B, Trinajstic N, Williams RW. 1993. Conformational preference functions for predicting helices in membrane proteins. *Biopolymers*, 33:255–73.

Karplus M, Petsko GA. 1990. Molecular dynamics simulations in biology. *Nature*, 347:631–9.

Kihara D, Kanehisa M. 2000. Tandem clusters of membrane proteins in complete genome sequences. *Genome Res*, 6:731–43.

Kihara D, Shimizu T, Kanehisa M. 1998. Prediction of membrane proteins based on classification of transmembrane segments. *Protein Engin*, 11:961–70.

Kim J, Moriyama EN, Coral GW, Peter JC, Carlson JR. 2000. Identification of novel multi-transmembrane proteins from genomic databases using quasi-periodic structural properties. *Bioinformatics*, 16:767–75.

Klein P, Kanehisa M, De Lisi C. 1985. The detection and classification of membrane-spanning proteins. *Biochim Biophys Acta*, 815:468–76.

Kleywegt G. 2001. ProSAL: Protein sequence analysis launcher [online]. Swedish Structural Biology Network. Accessed 15 February 2002. URL: http://alpha2.bmc.uu.se/~gerard/srf/prosal.html

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 305:567–80.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydrophathic character of a protein. *J Mol Biol*, 157:105–32.

Landau EM, Rosenbusch JP. 1996. Lipid cubic phases: a concept for the crystallization of membrane proteins. *Proc Natl Acad Sci USA*, 93:14532–5.

Le Borgne R, Hoflack B. 1998. Protein transport from the secretory to the endocytic pathway in mammalian cells. *Biochim Biophys Acta*, 1404:195–209.

Levitt M. 1983. Molecular dynamics of native proteins: I. Computer simulation of trajectories. *J Mol Biol*, 168:595–620.

Levitt M, Warshel A. 1975. Computer simulation of protein folding. *Nature*, 253:694–8.

Liu J, Rost B. 2001. Comparing function and structure between entire proteomes. *Protein Sci*, 10:1970–9.

Liu J, Tan H, Rost B. Eukaryotes full of loopy proteins? In possession of the author.

Mannella CA. 1998. Conformational changes in the mitochondrial channel protein, VDAC, and their functional implications. *J Struct Biol*, 121:207–18.

Marassi FM, Opella SJ. 2000. A solid-state NMR index of helical membrane protein structure and topology. *J Magn Reson*, 144:150–5.

Marchese A. 1999. Novel GPCRs and their endogenous ligands: expanding the boundaries of physiology and pharmacology. *Trends Pharmacol Sci*, 20:370–5.

Martin I, Dubois MC, Defrisep Quertain F, Saemark T, Burny A, Brasseur R, Ruysschaert JM. 1994. Correlation between fusogenicity and synthetic modified peptides corresponding to the NH2-terminal extremity of the SIV GP32 and their mode of insertion into the lipid bilayer; an infrared spectroscopy study. *J Virol*, 68:1139–48.

McDermott A, Polenova T, Bockmann A, Zilm KW, Paulson EK, Martin RW, Montelione GT, Paulsen EK. 2000. Partial NMR assignments for uniformly (13C, 15N)-enriched BPTI in the solid state. *J Biomol NMR*, 16:209–19.

McGovern K, Ehrmann M, Beckwith J. 1991. Decoding signals for membrane proteins using alkaline phosphatase fusions. *EMBO J*, 10:2773–82.

McGuigan JE. 1994. Antibodies to complementary peptides as probes for receptors. *Immunomethods*, 5:158–66.

Meyer JEW, Hofnung M, Schulz GE. 1997. Structure of Maltoporin from *Salmonella typhimurium* ligated with a Nitrophenyl-maltotrioside. *J Mol Biol*, 266:761–75.

Möller S, Croning DR, Apweiler R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, 17:646–53.

Möller S, Kriventseva EV, Apweiler R. 2000. A collection of well characterised integral membrane proteins. *Bioinformatics*, 16:1159–60.

Moreau JL, Huber G. 1999. Central adenosine A(2A) receptors: an overview. *Brain Res Brain Res Rev*, 31:65–82.

Morris GE, Sedgwick SG, Ellis JM, Pereboev A, Chamberlain JS, Nguyen thi M. 1998. An epitope structure for the C-terminal domain of dystrophin and utrophin. *Biochem*, 37:11117–27.

Nakai K, Horton P. 1999. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. *TIBS*, 24:34–6.

Nakai K, Kanehisa M. 1992. A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics*, 14:897–911.

Needleman SB, Wunsch CD. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, 48:443–53.

Ng P, Henikoff J, Henikoff S. 2000. PHAT: a transmembrane-specific substitution matrix. *Bioinformatics*, 16:760–6.

Nielsen H, Brunak S, von Heijne G. 1999. Machine learning approaches for the prediction of signal peptides and other protein sorting signals. *Protein Engin*, 12:3–9.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997a. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engin*, 10:1–6.

Nielsen H, Engelbrecht J, Brunak S, von Heijne G. 1997b. A neural network method for identification of prokaryotic and eukaroytoic signal peptides and prediction of their cleavage sites. *Int J Neural Syst*, 8:581–99.

Nielsen H, Engelbrecht J, von Heijne G, Brunak S. 1996. Defining a similarity threshold for a functional protein sequence pattern: the signal peptide cleavage site. *Proteins*, 24:165–77.

Nikaido H. 1994. Porins and specific diffusion channels in bacterial outer membranes. *J Biol Chem*, 269:3905–8.

Nilsson J, Persson B, von Heijne G. 2000. Consensus prediction of membrane protein topology. *FEBS Lett*, 486:267–9.

[NPSA] Network Protein Sequence Analysis [online]. 2001. Accessed 12 February 2002. URL: http://pbil.ibcp.fr/

Opella SJ, Ma C, Marassi FM. 2001. Nuclear magnetic resonance of membrane-associated peptides and proteins. *Meth Enzymol*, 339:285–313.

Pahl HL. 1999. Signal transduction from the endoplasmic reticulum to the cell nucleus. *Physiol Rev*, 79:683–701.

Pasquier C, Promponas VJ, Palaios GA, Hamodrakas JS, Hamodrakas SJ. 1999. A novel method for predicting transmembrane segments in proteins based on a statistical analysis of the SwissProt database: the PRED-TMR algorithm. *Protein Engin*, 12:381–85.

Paul C, Rosenbusch JP. 1985. Folding patterns of porin and bacteriorhodopsin. *EMBO J*, 4:1594–7.

Pautsch A, Schulz GE. 1998. Structure of the outer membrane protein A transmembrane protein. *Nat Struct Biol*, 5:1013–17.

Pebay-Peyroula E, Garavito RM, Rosenbusch JP, Zulauf M, Timmins PA. 1995. Detergent structure in tetragonal crystals of OmpF pori. *Structure*, 3:1051–9.

Pecheur EI, Sainte Marie J, Bienvenue A, Hoekstra D. 1999. Peptides and membrane fusion: towards an understanding of the molecular mechanism of protein induced fusion. *J Membr Biol*, 167: 1–17.

Persson B, Argos P. 1996. Topology prediction of membrane proteins. *Protein Sci*, 5:363–71.

Peuvot J, Schank A, Lins L, Brasseur R. 1999. Are the fusion processes involved in birth, life and death of the cell depending on tilted insertion of peptides into membranes? *J Theor Biol*, 198:173–81.

Promponas VJ, Palaios GA, Pasquier CM, Hamodrakas JS, Hamodrakas SJ. 1999. CoPreTHi: a Web tool which combines transmembrane protein segment prediction methods. *In Silico Biol*, 1:159–62.

Riek R, Pervushin K, Wuthrich K. 2000. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *TIBS*, 25:462–8.

Roberts MG, Phoenix DA, Pewsey AR. 1997. An algorithm for the detection of surface-active alpha-helices with the potential to anchor proteins at the membrane interface. *CABIOS*, 13:99–106.

Rost B. 1996. PHD: predicting one-dimensional protein structure by profile based neural networks. *Meth Enzymol*, 266:525–39.

Rost B. 1999. Twilight zone of protein sequence alignments. *Protein Engin*, 12:85–94.

Rost B. 2001. Protein secondary structure prediction continues to rise. *J Struct Biol*, 134:204–18.

Rost B, Casadio R, Fariselli P. 1996a. Refining neural network predictions for helical transmembrane proteins by dynamic programming. In States D, Agarwal P, Gaasterland T, Hunter L, Smith RF, eds. Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology (ISMB-96). 1996 Jun; St Louis, MO, USA. Menlo Park, CA: AAAI Pr. p 192–200.

Rost B, Casadio R, Fariselli P. 1996b. Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci*, 5:1704–18.

Rost B, Casadio R, Fariselli P, Sander C. 1995. Prediction of helical transmembrane segments at 95% accuracy. *Protein Sci*, 4:521–33.

Rost B, Sander C. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol*, 232:584–99.

Rost B, Sander C. 1994. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, 19:55–72.

Rost B, Sander C. 2000. Third generation prediction of secondary structure. *Methods Mol Biol*, 143:71–95.

Rost B, Sander C, Schneider R. 1994. PHD – an automatic server for protein secondary structure prediction. *CABIOS*, 10:53–60.

Rost B. 2000. PredictProtein [online]. Columbia University: New York. Accessed 12 February 2002. URL: http://cubic.bioc.columbia.edu/predictprotein

Rychlewski L. 2000. META server [online]. International Institute of Molecular and Cell Biology: Warsaw. Accessed 12 February. URL: www.bioinfo.pl/meta/; http://BioInfo.PL/LiveBench/

Sajot N, Genest M. 2000. Structure prediction of the dimeric neu/ErbB-2 transmembrane domain from multi-nanosecond molecular dynamics simulations. *Eur Biophys J*, 28:648–62.

Sanders CR, Nagy JK. 2000. Misfolding of membrane proteins in health and disease: the lady or the tiger? *Curr Opin Struct Biol*, 10:438–42.

Saragovi HU, Gehring K. 2000. Development of pharmacological agents for targeting neurotrophins and their receptors. *Trends Pharmacol Sci*, 21:93–8.

Schirmer T. 1998. General and specific porins from bacterial outer membranes. *J Struct Biol*, 121:101–9.

Schirmer T, Cowan SW. 1993. Prediction of membrane spanning beta-strands and its application to maltoporin. *Protein Sci*, 2:1361–3.

Schirmer T, Keller TA, Wang YF, Rosenbusch JP. 1995. Structural basis for sugar translocation through maltoporin channels at 3.1 A resolution. *Science*, 267:512–14.

Schulz GE. 1988. A critical evaluation of methods for prediction of protein secondary structures. *Annu Rev Biophys Biophys Chem*, 17:1–21.

Schulz GE. 2000. beta-Barrel membrane proteins. *Curr Opin Struct Biol*, 10:443–7.

Sedlacek HH. 2000. Kinase inhibitors in cancer therapy: a look ahead. *Drugs*, 59:435–76.

Sellers PH. 1974. On the theory and computation of evolutionary distances. *SIAM J Appl Math*, 26:787–93.

Seshadri K, Garemyr R, Wallin E, von Heijne G, Elofsson A. 1998. Architecture of beta-barrel membrane proteins: analysis of trimeric porins. *Protein Sci*, 7:2026–32.

Smith RF, Wiese BA, Wojzynski MK, Davison DB, Worley KC. 1996. BCM Search Launcher–an integrated interface to molecular biology data base search and analysis services available on the World Wide Web. *Genome Res*, 6:454–62.

Soltys BJ, Gupta RS. 2000. Mitochondrial proteins at unexpected cellular locations: export of proteins from mitochondria from an evolutionary perspective. *Int Rev Cytol*, 194:133–96.

Sonnhammer ELL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. In Glasgow J, Littlejohn T, Major F, Lathrop R, Sankoff D, Sensen C, eds. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology (ISMB-98). 1998 Jun 28–Jul 1; Montreal, Canada. Menlo Park, CA: AAAI Pr. p 175–82.

Stack JH, Horazdovsky B, Emr SD. 1995. Receptor-mediated protein sorting to the vacuole in yeast: roles for a protein kinase, a lipid kinase and GTP-binding proteins. *Annu Rev Cell Dev Biol*, 11:1–33.

Stadel JM, Wilson S, Bergsma DJ. 1997. Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery. *Trends Pharmacol Sci*, 22:162–5.

Tamm LK, Arora A, Kleinschmidt JH. 2001. Structure and assembly of beta-barrel membrane proteins. *J Biol Chem*, 276:32399–402.

Thanassi DG, Hutltgren SJ. 2000. Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol*, 12:420–30.

Traxler B, Boyd D, Beckwith J. 1993. The topological analysis of integral membrane proteins. *J Membr Biol*, 132:1–11.

Truscott KN, Pfanner N. 1999. Import of carrier proteins into mitochondria. *Biol Chem*, 380:1151–6.

Tusnady GE, Simon I. 1998. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J Mol Biol*, 283:489–506.

Tusnady GE, Simon I. 2001. Topology of membrane proteins. *J Chem Inf Comput Sci*, 41:364–8.

van Geest M, Lolkema JS. 2000. Membrane topology and insertion of membrane proteins: search for topogenic signals. *Microbiol Mol Biol Rev*, 64:13–33.

van Gunsteren WF. 1993. Molecular dynamics studies of proteins. *Curr Opin Struct Biol*, 3:167–74.

Vogel H, Jahnig F. 1986. Models for the structure of outer-membrane proteins of *Escherichia coli* derived from raman spectroscopy and prediction methods. *J Mol Biol*, 190:191–9.

von Heijne G. 1986. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J*, 5:3021–7.

von Heijne G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature*, 341:456–8.

von Heijne G. 1992. Membrane protein structure prediction. *J Mol Biol*, 225:487–94.

von Heijne G. 1996. Prediction of transmembrane protein topology. In Sternberg MJE, ed. Protein structure prediction. Oxford: Oxford Univ Pr. p 101–10.

Wallin E, von Heijne G. 1998. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci*, 7:1029–38.

Weiss MS, Schulz GE. 1992. Structure of porin refined at 1.8 Å resolution. *J Mol Biol*, 227:493–509.

Welte W, Weiss MS, Nestel U, Weckesser J, Schiltz E, Schulz GE. 1991. Prediction of the general structure of OmpF and PhoE from the sequence and structure of porin from *Rhodobacter capsulatus*: Orientation of porin in the membrane. *Biochim Biophys Acta*, 1080:271–4.

Wolfenden RV, Cullis PM, Southgate CCF. 1979. Water, protein folding, and the genetic code. *Science*, 206:575–7.

Wuthrich K. 2001. The way to NMR structures of proteins. *Nat Struct Biol*, 8:923–5.