Journal of Structural Biology 179 (2012) 327–337

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Membrane protein structural bioinformatics

Timothy Nugent, David T. Jones *

Bioinformatics Group, Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK

ARTICLE INFO

Article history:
Available online 31 October 2011

Keywords:
Membrane
Protein
Structure
Prediction
Re-entrant
Interfacial
Helix
Ab initio
Homology
Modelling
Molecular dynamics
Bioinformatics

ABSTRACT

Despite the increasing number of recently solved membrane protein structures, coverage of membrane protein fold space remains relatively sparse. This necessitates the use of computational strategies to investigate membrane protein structure, allowing us to further our understanding of how membrane proteins carry out their diverse range of functions, while aiding the development of novel predictive tools with which to probe uncharacterised folds. Analysis of known structures, the application of machine learning techniques, molecular dynamics simulations and protein structure prediction have enabled significant advances to be made in the field of membrane protein research. In this communication, the key bioinformatic methods that allow the characterisation of membrane proteins are reviewed, the tools available for the structural analysis of membrane proteins are presented and the contribution these tools have made to expanding our understanding of membrane protein structure, function and stability is discussed.

1. Introduction

Alpha-helical transmembrane (TM) proteins play a key role in a wide variety of essential biological processes including cell signalling, transport of membrane-impermeable molecules and intercellular communication, while many are also prime drug targets with some estimates suggesting that more than half of all drugs currently on the market target membrane proteins. Despite the biochemical and pharmacological importance of TM proteins, they remain significantly under-represented in structural databases. Since the first atomic-resolution TM protein structure was solved in 1985, progress has been slow with close to 300 unique structures deposited as of 2011 (Deisenhofer et al., 1985; Raman et al., 2006a; White, 2009). With advanced technologies such as synchrotron light sources, free-electron lasers and microfocus X-ray diffraction becoming available (Bowler et al., 2010; Bill et al., 2011), the number of structures will rise but in the meantime we rely on computational methods that allow us to extrapolate from the available experimental data. In this review article, we will focus on the analysis of known alpha-helical TM protein structures, highlighting research and tools that have helped to reveal the underlying principle of TM protein structure, function and stability. While beta-barrel TM proteins also perform a range of critical functions, their overall structural architecture is well conserved and shows limited diversity compared to the alpha-helical class. Tools to analyse beta-barrel TM proteins are generally restricted to topology predictors (Bagos et al., 2004; Fariselli et al., 2005; Bigelow and Rost, 2006; Ou et al., 2010; Freeman and Wimley, 2010; Singh et al., 2011), which are typically hidden Markov model-based, although some tools now allow contact and tertiary structure prediction (Randall et al., 2008). We will also look at how new developments in molecular dynamics simulations have allowed us to investigate the biological properties of TM proteins, investigate recent advances in three-dimensional structure prediction before finally highlighting a number of useful data resources.

2. Beyond the canonical topology concept

TM protein structure prediction has in recent years been dominated by topology prediction, where the total number of TM helices, their boundaries and in/out orientation relative to the membrane are detected using sequence-based methods. Early prediction methods were based on the physicochemical principle of a sliding window of hydrophobicity combined with von Heijne's 'positive-inside' rule (von Heijne, 1992), the observation that residues with positively charged side chains were clustered on intracellular loops. Such methods have since been superseded by machine learning approaches which utilise algorithms including hidden Markov models (HMMs), neural networks (NNs) and support vector machines (SVMs). These approaches prevail due to their ability

* Corresponding author. Fax: +44 (0)20 7387 1397.
E-mail addresses: t.nugent@cs.ucl.ac.uk (T. Nugent), d.jones@cs.ucl.ac.uk (D.T. Jones).

1047-8477/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2011.10.008

to learn complex relationships among the amino acids within a given window with which they are trained, particularly when provided with evolutionary information (Käll et al., 2005; Jones, 2007). Recent methods such as SPOCTOPUS and MEMSAT-SVM (Viklund et al., 2008; Nugent and Jones, 2009) that are trained using structurally derived topology data, rather than that acquired using various forms of biochemical characterisation, show particularly strong prediction performance with up to ~90% accuracy on certain data sets, highlighting the value of training such algorithms using high quality data. However, the SCAMPI method (Bernsel et al., 2008) achieves similar performance using an experimental scale of position-specific amino acid contributions to the free energy of membrane insertion, suggesting that prediction of TM protein topology from first principles is an attainable goal.

A number of topology prediction methods are also effective in identifying signal peptides, hydrophobic targeting motifs that are easily mis-predicted as TM helices. The first method to incorpotate signal peptide prediction was Phobius (Käll et al., 2004) which uses a HMM to successfully address the problem, while PolyPhobius (Käll et al., 2005) includes homology information to further increase prediction accuracy. Substructures such as re-entrant helices can also be identified by a number of predictors (see below). Other methods allow the incorporation of experimentally derived information in order to guide topology prediction, showing substantial benefits in prediction accuracy (Tusnady and Simon, 2001; Melen et al., 2003; Bagos et al., 2006).

Recently, more complex crystal structures have revealed a range of structural features that cannot be described by the standard topology model, including re-entrant helices, interfacial helices, kinks and coils within TM helices and tilted TM helices, while shedding light on TM helix packing and the function roles of specific residues within the membrane. Such proteins require us to redefine our canonical view of topology in order to accommodate these novel structural features.

2.1. Re-entrant helices

Re-entrant helices are short alpha-helices that enter and exit the membrane on the same side. A number of definitions exist that constitute re-entrant helices, ranging from parts of the sequence that penetrate the membrane to a depth of between 3 and 25 Å (Viklund et al., 2006) with the membrane located at 15 Å, to those that penetrate at least 6 Å but not more than 6 Å from the opposite face of the membrane (Nugent and Jones, 2009). Further subdivisions have been created including helix hairpins, which consist of two helices connected by a short coil region, helix followed by coil (or coil followed by helix) and finally re-entrant coil regions with irregular secondary structure. Re-entrant helices are common in many channel proteins and are believed to play an important role in gating, formation of channel selectivity filters and in pore formation. Such helices are typically less hydrophobic than true TM helices, and are frequently enriched in small amino acids such as alanine and glycine which are thought to assist the abrupt change in backbone direction that their structure requires. In aquaporin channels, re-entrant helices form a rigid selectivity filter with asparagine residues located on the re-entrant helices forming hydrogen bonds with water molecules that assist their single-file passage through the channel (Murata et al., 2000; Lee et al., 2005) (Fig. 1).

A number of TM topology predictors that now incorporate re-entrant helix prediction include TOP-MOD (Viklund et al., 2006), OCTOPUS (Viklund and Elofsson, 2008), SPOCTOPUS and MEMSAT-SVM. In a fully cross-validated test using a test set derived from crystal structures, MEMSAT-SVM correctly identified 44% (8 out of 18) re-entrant helices, with two false positive predictions, which compared favourably with OCTOPUS, a recent HMM and NN-based topology predictor that is also capable of predicting re-entrant helices. Whole proteome scans using MEMSAT-SVM suggests that 2–3% of both prokaryotic and eukaryotic TM proteins may contain re-entrant helices, although TOP-MOD, whose definition includes re-entrant regions that do not contain helices, estimates approximately 10–15%.

2.2. Interfacial helices

In the membrane–water interface region, a number of crystal structures show enrichment in interfacial alpha-helices which lie parallel to the membrane surface. Interface helices are located between 15 and 25 Å from the centre of the membrane and are conspicuous due to their enrichment in aromatic tryptophan and tyrosine residues compared to loop regions and TM helices (Granseth et al., 2005). Their side chains are typically orientated towards the membrane centre allowing the bulky, hydrophobic, six-membered rings to be buried inside the bilayer. Such helices are usually more hydrophobic than loop regions (Viklund et al., 2006) but less hydrophobic than TM or re-entrant helices, and may contain charged or polar residues which are orientated with their side chains facing away from the membrane (Fig. 2).

The functional roles of interface helices are poorly understood, but like re-entrant helices, are believed to regulate channel gating in both the KirBac 1.1 inward rectifying potassium channel (Doyle, 2004) and MscS mechanosensitive channel (Bass et al., 2003), while in photosystem I, interface helices appear to shield cofactors from the aqueous phase (Jordan et al., 2001). The observation that all TM helices separated by approximately 30 residues will have an interfacial helix between them suggests a more generalised role in constraining inter-helix distances and therefore reducing structural freedom (Granseth et al., 2005).

Currently, two methods exist that can predict interfacial helices from sequence. TOP-MOD (Viklund et al., 2006) is capable of predicting interfacial helices in addition to TM topology, by modelling interface helices as one of four structural states within a HMM. TOP-MOD was able to detect 42% of interface helices with 75% specificity under cross validation using a data set of 79 sequences derived from crystal structures. Another method uses a variant of the hydrophobic moment analysis as part of a protocol for distinguishing interface helices from TM or extramembranous helices with accuracy of 72% (Orgel, 2004).

2.3. Kinks and coils in the membrane region

A number of recent studies have investigated the presence of kinks and coils within TM helices that disrupt the helical backbone and lead to a deviation in helix direction. Kauko et al. analysed residues deep within the membrane that were in a coil state, observing that approximately 7% of residues in the membrane core formed coils, and that such regions were found in TM helices as both kinks and major breaks in helix structure, as well as in parts of re-entrant helices (Kauko et al., 2008). Additionally, they noted that residues in such regions frequently contained polar side chains, and were therefore either buried or located adjacent to an aqueous channel. They were also significantly more conserved than other residues in the membrane. The frequency at which they were found in channel and transporter proteins suggests they play a role in flexibility and in introducing the necessary polarity required for transport across the membrane, thus are essential for the function of such proteins (Fig. 3).

Hall et al. analysed a database of TM protein crystal structures and showed that 44% of TM helices contained a significant helical kink, with proline causing 35% of these (Hall et al., 2009). The distinctive cyclic structure of proline's side chain locks its φ

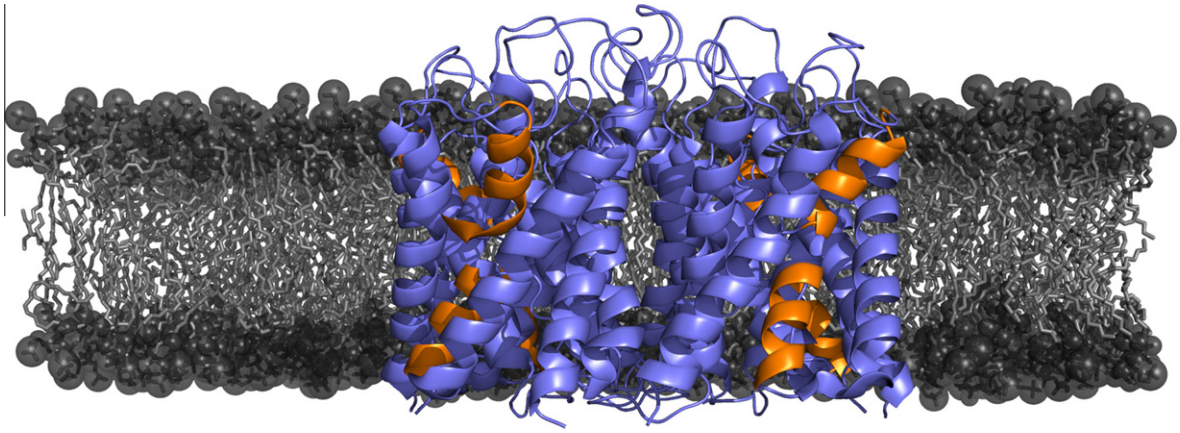

Fig. 1. Two chains from aquaporin-1 (PDB code 1h6i) with re-entrant helices shown in orange. Each chain of the homotetrameric channel protein contains six transmembrane and two re-entrant helices.

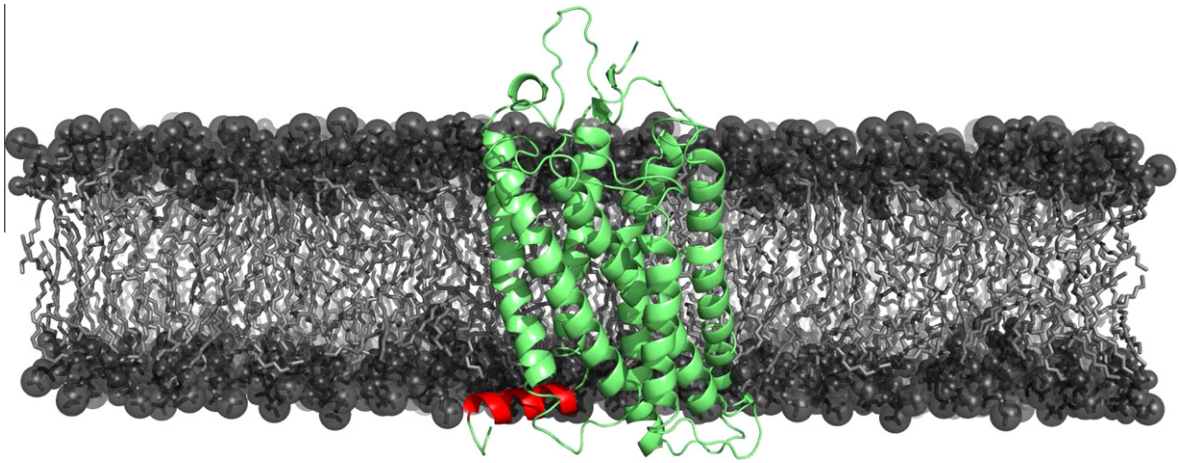

Fig. 2. Rhodopsin structure (PDB code 1gzm) with an interfacial helix shown in red. Evidence suggests that, in addition to its role in binding the G protein transducin, the interfacial helix acts as a membrane-dependent conformational switch domain which adopts a helical structure only in the presence of membranes (Krishna et al., 2002).

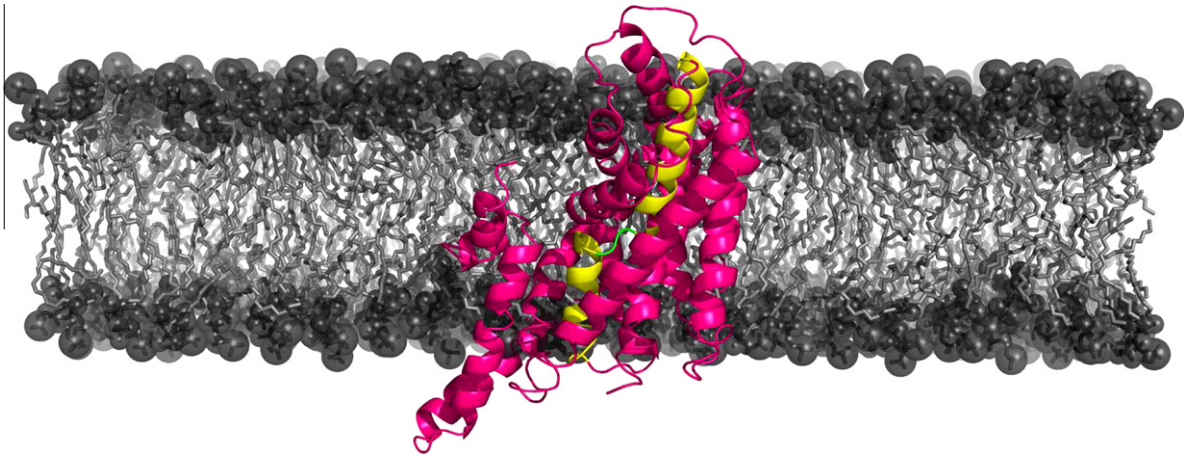

Fig. 3. Single chain from the homotrimeric proton glutamate symport protein (PDB code 2nwl), showing a coil region (green) breaking a transmembrane helix (yellow). The coil segment is likely to form a flexible binding site during transport allowing the large conformational change that is required (Kauko et al., 2008).

backbone dihedral angle at approximately −75°, inducing the helical kink. It was also shown that many non-proline helical kinks could be caused by residues such as serine and glycine when located at the centre of a helical kinks. The side chain of serine forms a hydrogen bond with the main chain carbonyl of the i −4th or i +4th residue, thus creating the kink.

Another study proposed an evolutionary hypothesis in which a mutation to proline initially induces the kink (Yohannan et al., 2004). The resulting packing defects were later repaired by further mutation, locking the kink into the structure. Subsequent removal of proline residues did not therefore have major structural consequences. Evidence of such *vestigial prolines* can be found by examining the corresponding positions in homologous proteins. Consistent with this hypothesis, prolines were found in homologous sequences for 14 out of 17 non-proline kinks in TM proteins of known structure.

Predictions of helical kinks can be made using a multiple sequence alignment, with the method of Yohannan et al. (2004) able to predict kink positions with >90% reliability. A recent approach, MC-HELAN, uses a Monte Carlo-based algorithm to determine helical axes alongside positions and angles of helical kinks (Langelaan et al., 2010).

2.4. Tilted transmembrane helices

The topological depiction of TM proteins as a series of ideal helices packed in a tight bundle has been challenged by the wide distribution of helix tilt angles with respect to the membrane normal that can be observed in crystal structures. Tilting of TM helices is thought to be the major response to hydrophobic mismatch – the difference between the hydrophobic length of a TM helix and the hydrophobic thickness of the membrane it spans (Duque et al., 2002; Bowie, 2005; Kim and Im, 2010) – although the presence of charged or aromatic anchor residues also appears to play an important role (Vostrikov et al., 2010).

Kim and Im performed sampling simulations and calculated the potentials of mean force as a function of TM helix tilt angle under various hydrophobic mismatch conditions. They noted that tilting of TM helices by up to $\sim 10^{\circ}$ was inherent due to the intrinsic entropic contribution arising from helix precession around the membrane normal, even under a negative mismatch (where the TM helix is shorter than the hydrophobic thickness of the membrane), since the accessible orientational space of the helix is reduced as the tilt angle decreases. Under a positive mismatch, favourable helix–lipid interactions are believed to provide the additional driving forces for TM helix tilting $>10^{\circ}$, while minimum tilting generally occurs where there is hydrophobic match with little lipid perturbation.

Anchoring residues at the hydrophilic/hydrophobic interface were also shown to be an important determinant of TM helix orientation. Arginine, lysine, and tryptophan residues typically showed preferences for positions at this interface, with helices at the minimum tilt angle, regardless of mismatch. Vostrikov et al. (2010) generated peptides containing leucine–alanine spacers between paired tryptophan residues, with lysine, arginine or glycine anchors, before using solid-state ^{2}H NMR spectroscopy to investigate helix orientations in phosphatidylcholine bilayers of varying thickness. When the anchor residue was lysine, arginine, or glycine, the direction of the tilt was essentially constant in different lipids and presumably is dictated by the tryptophan residues flanking the inner helical core. When the anchoring residues were replaced with tryptophan, the tilt direction changed dramatically and seemed to introduce additional dynamics while flattening the dependence of the tilt magnitude on bilayer thickness. It appears that interfacial indole rings play an important role in determining helix tilt by accommodating lipid chains in a groove formed by the indole ring and a series of short side chain residues (Beswick et al., 2011), but that competition between multiple indole groups may have a destabilising effect.

Few computational approaches exists for the predicting of TM helix tilt angles which is surprising since such a measure would seem a good target for a regression model. To date, molecular dynamics studies have been used to accurately predict tilt angles (Sankararamakrishnan and Weinstein, 2000) while another method, which uses rotamer statistics to infer typical side chain positions and compute local tilt angles, shows good agreement with hydrophobicity-based approaches (Cheng et al., 2004).

2.5. Polar residues in the membrane

While the presence of proline residues in the membrane can be explained by their necessary kink-forming propensity, the occurrence of other polar amino acids within the lipid bilayer is harder to explain due to the energetic difficulties in accommodating their side chains in a predominantly hydrophobic environment. Polar residues are found within the membrane infrequently since contact with hydrophobic lipid tails contributes negatively to overall stability (White and Wimley, 1999). The majority are therefore either buried within the interior of the protein, and thus shielded from lipids, or exposed to an internal water-filled cavity while the formation of hydrogen bonds or salt bridges with other polar residues may stabilise the remainder. Since most salt bridges in TM proteins occur between helices rather than within a single helix, the formation of such interactions favours topologies with large numbers of helices (Jayasinghe et al., 2001a). Additionally, single TM helices that contain polar residues are unfavourable for membrane insertion by the translocon machinery (Hessa et al., 2007). It is also common for polar side chains to snorkel out towards the phospholipid head groups, and in some cases induce a local deformation of the membrane (Granseth et al., 2005; Wee et al., 2011).

A recent study of the functional roles played by polar residues in known TM protein structures indicates they are frequently involved in the binding of small compounds, particularly in transporters and channels (Illergard et al., 2011), while other functions include gating (Yamashita et al., 2005), flexibility (Abramson et al., 2003), and proton transfer (Seeger et al., 2006). Multiple sequence alignments revealed that polar residues within the membrane are more highly conserved than both non-polar residues within the membrane and polar residues outside the membrane, most likely because they play a key functional role and because they are buried in the protein interior. Of particular functional importance were histidine residues which often interact with prosthetic groups in photosynthetic and oxidoreductase-related proteins (Lancaster et al., 1999; Huang et al., 2006).

A method capable of identifying polar residues within the membrane directly from sequence was also presented. Applying this method to the human proteome indicated that polar residues occurred most frequently in GPCR and active transporter proteins, and least frequently in protein families with few TM helices (Illergard et al., 2011). Two methods, HOLLOW (Ho and Gruswitz, 2008) and Pore-Walker (Pellegrini-Calace et al., 2009), exist that are able to analyse channels and interior surfaces in TM protein structures, where polar residues are frequently found. These automated methods identify the pore centre and pore axis using geometric criteria, allowing the biggest and longest cavity through the channel to be detected. Pore features, including diameter profiles, pore-lining residues, size, shape and regularity can then be calculated.

2.6. Helix–helix interactions

Interactions between TM helices are important determinants for TM protein folding, function, dynamics and even structural classification (Fuchs and Frishman, 2010). A number of studies have identified structural and sequence motifs that recur frequently in helix–helix interaction sites in TM proteins. Walters and DeGrado analysed 445 interacting helical pairs from 32 high resolution structures, clustering them according to their three-dimensional similarity so that each cluster member was within

1.5 Å C-alpha RMSD of a reference centroid pair. They found that 29% of pairs fell into the most populous cluster, 74% of pairs fell into the top five most populous clusters, and 90% fell within the top 14. The largest of these consisted of an anti-parallel motif with left-handed packing angles, stabilised by the packing of small side chains every seven residues, while right-handed parallel and anti-parallel structures showed a similar tendency though spaced at four-residue intervals (Walters and DeGrado, 2006). Harrington and Ben-Tal, 2009 showed that the packing interactions of TM helices from 15 diverse proteins could be classified into five kinds of specific favourable interhelical interactions – two types of packing motif, hydrogen bonds, aromatic interactions and salt bridges (Harrington and Ben-Tal, 2009) (Fig. 4).

Packing motifs are typically composed of close knob-in-hole type packed residues with side chains of limited conformational flexibility, allowing each residue to fill a cavity and make van der Waals (and occasionally polar) contacts without significant entropic losses. The two types of packing interaction were either small residues (G/A/S/C) as knobs or I/V/L/T residues as knobs in I/V/L/T contact patches, where at least one of the surrounding hole residues must be I/V/L/T. The most commonly studied motifs include the GXXXG motif found in glycophorin A (Lemmon et al., 1992; Schneider and Engelman, 2004) and heptad motifs of leucine residues (Gurezka et al., 1999), while less frequent ones involved proline packing (Senes et al., 2004).

Hydrogen bonds between TM helices are on average weaker than those found in globular proteins but contributed significantly to stability, particularly between polar residues (Zhou et al., 2001; Gratkowski et al., 2001), although their strength varied significantly according to environment (Joh et al., 2008). Participating residues were found to be highly conserved (Hildebrand et al., 2008). Another study identified a specific aromatic pattern, aromatic-XX-aromatic, which was demonstrated to stabilise helix–helix interactions during assembly (Sal-Man et al., 2007). Aromatic interactions include cation–π interactions, a non-covalent molecular interaction between the face of an electron-rich π system (such as an aromatic rings) with an adjacent cation (Dougherty, 1996; Shi et al., 2002), aromatic stacking (Johnson et al., 2007) and interactions between polar atoms and aromatic rings, where the edges act as weak electron donors and the centres as weak acceptors (Nanda and Schmiedekamp, 2008). As outlined earlier, salt bridges occur rarely in TM proteins (Fig. 5).

Prediction of helix–helix interactions from sequence has only recently been investigated in TM proteins due to the relative lack of TM protein crystal structures. In contrast, a range of tools exist to predict residue contacts in globular proteins (Punta and Rost, 2005; Cheng and Baldi, 2007), and such methods have even been used to assess globular protein models submitted to the Critical Assessment of Structure Prediction (CASP) experiment (Izarzugaza et al., 2007). Unfortunately, such globular proteins contact predictors perform poorly when applied to TM proteins, most likely due to the differences between TM and globular interaction motifs (Fuchs et al., 2009). Recently, a number of machine learning-based methods have been developed that attempt to predict residue contacts and helix–helix interaction in TM proteins, trained using structural data. TMHcon implements a neural network in combination with profile data, residue co-evolution information, predicted lipid exposure using the LIPS method (Adamian and Liang, 2006), and a number of TM protein specific features in order to predict helix–helix interaction (Fuchs et al., 2009). TMhit uses a two-level hierarchical approach in combination with a SVM classifier, firstly discriminating between contacts and non-contacts on a per residue basis, before determining the structure of the contact map from all possible pairs of predicted contact residues and therefore avoiding the high computational cost incurred by the quadratic growth of residue pair prediction (Lo et al., 2009). MEMPACK (Nugent and Jones, 2010) also uses SVM classifiers to predict lipid exposure, residue contacts, helix–helix interactions and finally the optimal helical packing arrangements of TM proteins. An SVM trained using molecular dynamics (MD) data initially identifies residues potentially exposed to lipid, before incorporating this information with PSI-BLAST (Altschul et al., 1997) profile data and a variety of sequence-based features to train an additional SVM to predict residue contacts. Combining these results with *a priori* topology information, prediction of helix–helix interaction was possible with up to 65% accuracy under stringent cross-validation on a non-redundant test set of 74 protein chains.

3. Molecular dynamics simulations of membrane proteins

Due to advances in atomistic (AT) simulation algorithms, computer hardware, the introduction of coarse-grained (CG) models, and new high-resolution crystal structures, MD simulations of membrane proteins have made significant progress in recent years. Calculations incorporating proteins, lipids, water and ions typically involve between 50,000 atoms for small proteins, and up to 300,000 atoms for the largest proteins yet studied, therefore posing a major computational challenge (Gumbart et al., 2005). Despite this, it has also been possible to simulate a number of transporters including potassium, aquaporin and mechanosensitive channels (Fig. 6) as well as β-barrel outer membrane proteins (Roux and Schulten, 2004), in addition to investigating the interactions of

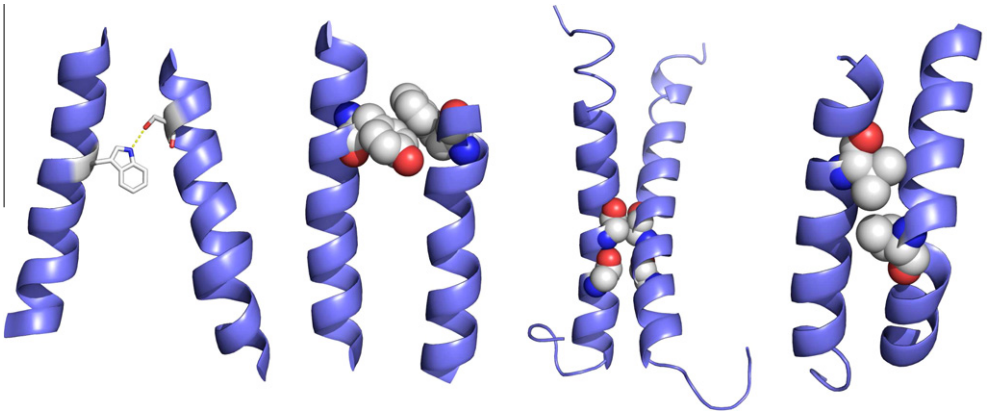

Fig. 4. Common helix–helix interaction types in transmembrane helices. (a) Hydrogen bond between tryptophan and serine residues in helices from the β-adrenergic GPCR (PDB code 2rh1). (b) Aromatic interactions between tryptophan residues in sensory rhodopsin (PDB code 1xio). (c) GXXXG motif interaction in glycophorin A (PDB code 1afo). (d) Valine residues form a knob in contact patch interaction in the mitochondrial ADP/ATP carrier (PDB code 1okc).

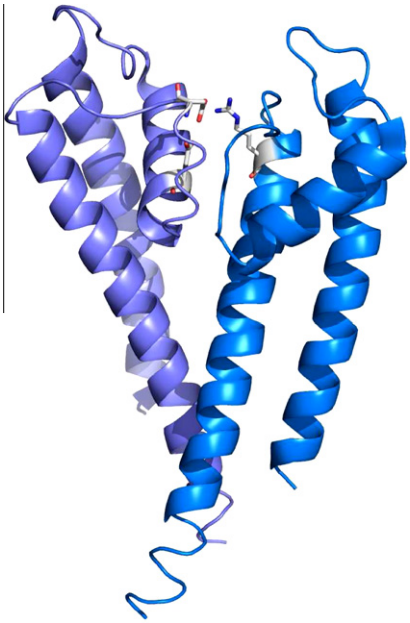

Fig. 5. Salt bridges have been shown to form between Arg89 of one strand and Asp80 and Glu71 (located on the re-entrant region) of the adjacent strand in models of the KcsA potassium channel (PDB code 1jvm) (Stewart, 2009). Interacting residues are shown in stick rendering.

individual amino acids within the bilayer. Such studies have provided insight into the physical mechanisms underlying membrane processes, and demonstrated that membranes behave adaptively rather than being purely as hydrophobic solvents, with important functional implications for TM protein interactions (Lindahl and Sansom, 2008).

3.1. Atomistic (AT) simulations

Because of the large size of membrane systems and the required timescales, AT-detail studies generally fall short of allowing biological properties to be predicted. However, several recent large-scale AT simulations of complex TM proteins have been reported. A significant amount of effort has been directed towards understanding the voltage sensing (VS) mechanism of the kV channel, with AT studies having focused on both intact kV tetramers and isolated VS domains. While typical TM helices are composed of a core of hydrophobic residues with charged or aromatic residues at each end forming interactions with lipid headgroups or water, the S4 helix of the kV channel contains four or five basic residues, mostly arginines, along its length yet is able to adopt a TM orientation as part of an independently stable VS domain. AT simulations reveal that the charged S4 helix is accommodated through a combination of hydrogen bonding to other parts of the protein and lipid carbonyl/phosphate groups, as well as penetration of water molecules deep into cavities on both sides of the membrane (Freites et al., 2006; Sands and Sansom, 2007). A recent multiscale MD study also demonstrated that the preference of the hydrophobic residues of the S4 helix to enter the membrane dominated the free-energy penalty for insertion of charged residues, and was accompanied by local distortion of the bilayer (Wee et al., 2011). Another AT study on the kV channel focused on the re-entrant helix that forms the pore, demonstrating that the tertiary re-entrant fold is present in kV monomers and is established early in biogenesis and independently of subunit tetramerisation (Gajewski et al., 2011).

Channel gating of the MscS and MscL mechanosensitive channels, members of a family of bacterial TM proteins that function as a safety valve to relieve the turgor pressure produced by osmotic downshock, have also been investigated using AT simulations. Akitake et al. explored the pathway for conformational transitions in MscS, observing complete straightening of the third TM helix in the open state, and buckling of the helix at two different sites upon return to the compact, non-conductive conformation (Akitake et al., 2007). Another study observed that MscL channel opening was possible only following application of a radial bias force to each of the five subunits of MscL in an outward direction (Jeon and Voth, 2008). A simulation focusing on the SecYEβ protein conducting translocon revealed the formation of a plug which swings open during translocation and closes thereafter, while a ring region forms a tight, elastic seal around the translocating oligopeptides that blocks water and ions in the closed state (Gumbart and Schulten, 2006).

3.2. Coarse-grained (CG) models

More recently, CG simulations have been used to study membrane systems. CG models treat small groups of 3–5 heavy atoms

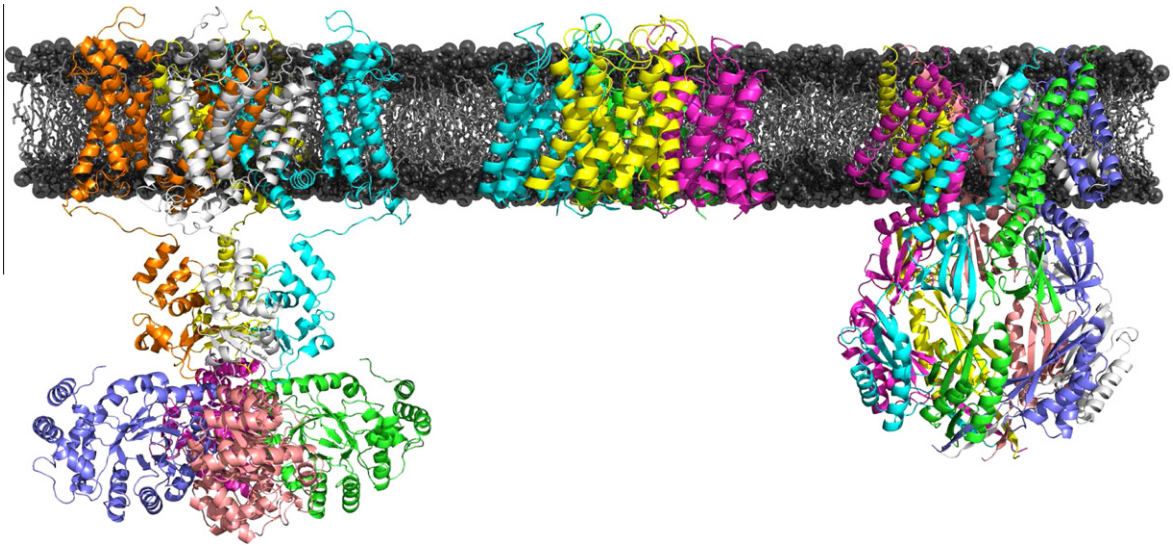

Fig. 6. Transmembrane channel proteins that have been studied using molecular dynamics. From left to right: kV potassium channel (PDB code 2r9r), aquaporin-1 (PDB code 1h6i) and the MscS mechanosensitive channel protein (PDB code 2vv5).

as single particles, enabling extended (>100 ns) timescales to be assessed. This representation leads to an order of magnitude fewer interactions allowing speedups of 2–3 orders of magnitude compared to AT simulations (Marrink et al., 2007). Bond et al. used CG–MD to successfully simulate the insertion of synthetic model membrane peptides into a lipid bilayer, in agreement with experimental biophysical data, before extending the approach to a more complex TM protein, the bacterial sugar transporter LacY. Comparison of a 200 ns CG–MD simulation of LacY in a DPPC bilayer with a 50 ns atomistic simulation of the same protein in a DMPC bilayer revealed that the two methods yielded comparable predictions of lipid–protein interactions (Bond et al., 2007). CG–MD has also been used to study kV and mechanosensitive channels at microsecond timescales, revealing possible gating mechanisms that would be beyond the reach AT simulations therefore demonstrating the utility of CG–MD for studies of membrane systems (Yefimov et al., 2008; Treptow et al., 2008).

MD simulations have proved effective for investigating protein localisation within membranes. Scott et al. used CG–MD to self-assemble lipid bilayers around TM proteins, allowing the position and orientation of ~100 non-redundant structures to be determined, revealing how local bilayer deformations were related to different TM protein class and lipid species (Scott et al., 2008). Periole et al., 2007 used CG–MD to simulate 16 rhodopsin molecules within a single membrane system, also noting local membrane deformation and suggesting this may play a key role in defining the rate, extent, and orientational preference of protein–protein association (Periole et al., 2007). Simulations have also explored bacterial β-barrel outer membrane proteins such as OmpA and OpcA, examining the conformational dynamics when transplanted from a crystal to bilayer environment (Bond et al., 2006), as well as side chain interactions with lipids, exploring dehydration, solvation and partition free energies (Johansson and Lindahl, 2006; MacCallum et al., 2008; Friemann et al., 2009). Specific residues known to play important structural or functional roles such as tryptophan and arginine have been studied in depth (Norman and Nymeyer, 2006; Dorairaj and Allen, 2007; Gumbart et al., 2011; Schow et al., 2011), as has ion selectivity by membrane transport proteins (Roux, 2010; Yu et al., 2010; Roux et al., 2011).

4. Three-dimensional structure prediction

As with globular proteins, three-dimensional structure prediction of TM proteins can be dealt with via two strategies, *ab initio* and homology modelling.

4.1. Ab initio modelling

Ab initio (or *de novo*) modelling approaches attempt to build three-dimensional models based on physical principles rather than using known crystal structures evolutionary related to the target protein. This typically involves recombining short segments of the chain preset to different low-energy local conformations, while searching for the global minimum using a suitable energy function and Monte Carlo methods. Other approaches make use of alternate lower-resolution representations of proteins and other sampling techniques. While the majority of *ab initio* prediction methods focus on globular proteins, some have been adapted to deal specifically with TM protein. Such efforts are particularly important considering the relatively low coverage of TM protein fold space compared to their globular counterparts.

FILM (Pellegrini-Calace et al., 2003) is a modification of the FRAGFOLD (Jones, 1997, 2001) globular protein structure prediction method, based on the assembly of supersecondary structural fragments using a simulated annealing algorithm. FRAGFOLD attempts to reduce conformational search space by preselecting fragments from a library of highly resolved protein structures. FILM added a membrane potential to the FRAGFOLD energy terms (pairwise, solvation, steric and hydrogen bonding) which was derived from the statistical analysis of a data set comprising 640 TM helices with experimentally defined topology. The membrane is modelled implicitly as an anisotropic planar slab composed of several layers, each representing a region of the membrane where the physicochemical properties influence the amino acid composition of the protein. When tested on five small membrane-embedded sequences each composed of up to 79 residues and two TM helices, FILM was able to predict the correct topology and conformation of all but one of the peptides, at a reasonable level of accuracy.

Another *ab initio* globular protein prediction method – Rosetta (Simons et al., 1997; Simons et al., 1999; Rohl et al., 2004) – was also modified to deal specifically with TM proteins. Rosetta uses potential functions for computing the lowest energy structure of an amino acid sequence, while continually improving the potential functions and search algorithms with feedback from predictions and is frequently ranked among the top performing *ab initio* methods at CASP experiments (Ben-David et al., 2009). RosettaMembrane (Barth et al., 2007) implements an energy function that describes membrane intraprotein interactions at atomic level and membrane protein/lipid interactions implicitly, while treating hydrogen bonds explicitly. RosettaMembrane was able to predict the structures of 12 small TM protein domain (<150 residues) successfully to within 4 Å RMSD of the native structure, suggesting that the model captures the essential physical properties that govern the solvation and stability of membrane proteins and compares favourably with predictions obtained on small water-soluble protein domains. The RosettaMembrane method has also been combined with homology modelling and domain assembly methods to model the open and closed states of the kV and KvAP potassium channels, providing insight into the mechanism of voltage-dependent gating through conformational change (Yarov-Yarovoy et al., 2006).

4.2. Constrained prediction

RosettaMembrane was subsequently modified to allow constraints to be imposed on helix–helix packing arrangements at particular positions predicted from sequence or identified by experiments. This allowed larger (between 90 and 300 residues) structures with diverse functions and topologies to be predicted, with results suggesting that the population of near-native models could be enriched by enforcing only a single constraint during folding simulations in 9 of the 12 TM proteins tested, while models within 4 Å of the native structure could be achieved in four cases (Barth et al., 2009). While the method does not perform contact prediction itself, it demonstrates that near native structure can be achieved for complex membrane proteins if even limited residue–residue interaction data is available. As discussed earlier, such information is available via a number of TM-specific methods (Fuchs et al., 2009; Lo et al., 2009; Nugent and Jones, 2010) while approaches that are able to predict lipid exposure within the membrane are also likely to augment *ab initio* folding (Adamian and Liang, 2006; Illergard et al., 2010; Nugent and Jones, 2010).

A number of methods have attempted the *ab initio* prediction of GPCR structures (Becker et al., 2004; Freddolino et al., 2004; Li and Goddard, 2008), producing accurate models using a variety of strategies including CG and AT model refinement using MD simulations, while other approaches deal specifically with extramembranous loop regions (Hildebrand et al., 2009; Goldfeld et al., 2011). Recently, the first blind prediction of a GPCR target was held for the adenosine A_{2A} receptor with results indicating that accurate prediction of GPCR structure and ligand interactions remains a

challenge. The most successful prediction methods relied on homology modelling approaches based on the template structures of β-adrenergic receptors, although extracellular loop regions were predicted *ab initio* in a number of the top predictions (Michino et al., 2008).

4.3. Homology modelling

Homology (or comparative) modelling involves the use of related template structures in order to build a three-dimensional model of the target protein, based on the observation that protein structure is conserved more highly than amino acid sequence hence proteins that have diverged in sequence but still share detectable similarity may still share common structural properties, and particularly the overall fold. As with *ab initio* modelling methods, very few homology modelling tools deal specifically with TM proteins, although research has demonstrated that tools designed for globular proteins perform at least as well on TM proteins (Forrest et al., 2006). Recently, a TM-specific homology-based tool MEDELLER was demonstrated to outperform the popular homology modelling tool MODELLER (Eswar et al., 2006) when tested using TM proteins. Using 616 target-template pairs of TM proteins, MEDELLER gave an average backbone root mean square deviation (RMSD) of 2.6 versus 3.2 Å for Modeller (Kelm et al., 2010). Other methods deal specifically with GPCRs, such as FoldGPCR which uses predicted conserved residue-residue contacts between the template and target while exploiting an all-atom implicit membrane force field (Michino et al., 2010). The TASSER method, at the interface of *ab initio* and homology modelling, works by threading the query sequence onto parts of solved proteins, before refining the resulting structural template. Application to all human GPCRs yielded 12 structures with RMSD within 4 Å of native, although many others above 6 Å (Zhang et al., 2006).

4.4. Model quality assessment

Model quality assessment plays a crucial role in protein structure prediction, allowing high-quality models to be discriminated from erroneous ones by attempting to correlate a score with a measure of model quality, as measured by the distance to the native structure. Standard methods for measuring model quality such as Procheck (Laskowski et al., 1993) and WhatIf (Vriend, 1990) assess local stereochemistry but do not consider the global fold of a model, while more recent methods that measure absolute quality were shown to perform poorly on TM structures (Benkert et al., 2011). Recently, a scoring function trained on globular proteins, ProQres, was re-trained using a SVM and structural models of TM proteins, with the addition of TM protein-specific features such as topology and Z-coordinate predictions (Wallner and Elofsson (2006); Ray et al., 2010). When applied to various TM protein test sets, ProQM was shown to outperform all methods developed for globular proteins while maintaining good performance on extramembranous domains. It was also able to accurately predict GPCR quality at local and global levels.

MD simulations have also been used to assess TM protein model quality. A comparison between a MsbA model simulated in a lipid bilayer with another bacterial ABC transporter, BtuCD, showed the latter structure demonstrated good conformational stability in the same bilayer environment over the same timescale. Comparative analyses of the two simulations revealed significant differences, mainly due to dramatic structural deformations of MsbA, suggesting that MD could be used effectively to measure TM protein model quality, by allowing the global conformational stability of a model to be assessed (Ivetac and Sansom, 2008).

5. Databases

A number of databases serve as repositories for the sequences and structures of TM proteins. PDB_TM (Tusnady et al., 2005a), OPM (Lomize et al., 2006b), MPDB (Raman et al., 2006b) and CGDB (Chetwynd et al., 2008) all contain TM proteins of known structure determined using a variety of experimental techniques. PDB_TM, OPM, and CGDB additionally contain orientation predictions of the protein relative to the membrane based on hydrophobicity/structural feature analysis (Tusnady et al., 2005b), water–lipid transfer energy minimisation (Lomize et al., 2006a) and CG–MD simulations (Sansom et al., 2008). OPM also provides N-terminus localisation information, useful for topological studies, while TOPDB (Tusnady et al., 2008) and MPtopo (Jayasinghe et al., 2001b) include TM proteins of unknown 3D structure whose topologies have been experimentally validated using low-resolution biochemical techniques such as gene fusion, antibody and mutagenesis studies. TMPad (Lo et al., 2011), a repository of helix–helix interactions in TM proteins, is likely to be useful for contact studies and Lipidbook (Domanski et al., 2010), a repository for force field parameters used in membrane simulations, provides a MD resource. A number of TM protein databases collect information on specific families including voltage-gated potassium channels (Gallin and Boutet, 2011) and GPCRs (Vroling et al., 2011), while others such as LGICdb (Donizelli et al., 2006) and TCDB (Saier et al., 2006) focus on particular structural or functional classes. The CATH (Class, Architecture, Topology, Homology) database (Cuff et al., 2011), a manually curated classification of protein domain structures, has recently been extended to include TM proteins. Structural classification of TM proteins is more difficult than for soluble proteins, partly due to the limited number of structural arrangements, and also their tendency to be structurally similar, regardless of evolutionary history or function (Neumann et al., 2010). CATH 3.4 includes 2274 new TM proteins, accounting for 71 new superfamilies and 22 new fold groups. Sixty-two percent of the newly classified superfamilies are alpha-helical in nature, while 24% are single transmembrane helix superfamilies.

6. Concluding remarks

Recent analysis of existing TM protein structures has revealed a range of diverse structural features that force us to redefine our traditional concept of topology, and this will surely continue as deposition of TM protein structures in database approaches exponential rates and novel topological complexity is revealed. This will undoubtedly assist three-dimensional structure prediction, where steady progress is being made. However, despite the promising results obtained using *ab initio* structure prediction methods, a number of issues need to be addressed before we are able to fold large TM proteins *in silico*. Enhanced conformational sampling strategies and better treatment of electrostatics might be used to overcome current limitations, while the use of a more restricted supersecondary structure fragment library, perhaps based solely on TM protein structures, could aid fragment based approaches. Using predicted or experimental data to constrain predictions may further help to reduce the combinatorial complexity of such methods. Although focused on GPCRs and limited by the need for suitable templates, homology modelling approaches continue to make advances and initiatives such as the Critical Assessment of GPCR Structure Modelling and Docking will help to push the field forwards. Molecular dynamic simulations of TM proteins are also making significant advances, with simulations now approaching microsecond timescales even for complex structures, particularly via the use of CG approaches. Such developments now allow novel and quantitative predictions to be made about biological properties that may help direct further experimental work.

References

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H.R., Iwata, S., 2003. Structure and mechanism of the lactose permease of *Escherichia coli*. Science 301 (5633), 610–615.

Adamian, L., Liang, J., 2006. Prediction of transmembrane helix orientation in polytopic membrane proteins. BMC Struct. Biol. 6, 13.

Akitake, B., Anishkin, A., Liu, N., Sukharev, S., 2007. Straightening and sequential buckling of the pore-lining helices define the gating cycle of mscs. Nat. Struct. Mol. Biol. 14 (12), 1141–1149.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17), 3389–3402.

Bagos, P.G., Liakopoulos, T.D., Hamodrakas, S.J., 2006. Algorithms for incorporating prior topological information in HMMs: application to transmembrane proteins. BMC Bioinf. 7, 189.

Bagos, P.G., Liakopoulos, T.D., Spyropoulos, I.C., Hamodrakas, S.J., 2004. A hidden markov model method, capable of predicting and discriminating beta-barrel outer membrane proteins. BMC Bioinf. 5, 29.

Barth, P., Schonbrun, J., Baker, D., 2007. Toward high-resolution prediction and design of transmembrane helical protein structures. Proc. Natl. Acad. Sci. U S A 104 (40), 15682–15687.

Barth, P., Wallner, B., Baker, D., 2009. Prediction of membrane protein structures with complex topologies using limited constraints. Proc. Natl. Acad. Sci. U S A 106 (5), 1409–1414.

Bass, R.B., Locher, K.P., Borths, E., Poon, Y., Strop, P., Lee, A., Rees, D.C., 2003. The structures of btucd and mscs and their implications for transporter and channel function. FEBS Lett. 555 (1), 111–115.

Becker, O.M., Marantz, Y., Shacham, S., Inbal, B., Heifetz, A., Kalid, O., Bar-Haim, S., Warshaviak, D., Fichman, M., Noiman, S., 2004. G protein-coupled receptors: in silico drug discovery in 3d. Proc. Natl. Acad. Sci. U S A 101 (31), 11304–11309.

Ben-David, M., Noivirt-Brik, O., Paz, A., Prilusky, J., Sussman, J.L., Levy, Y., 2009. Assessment of casp8 structure predictions for template free targets. Proteins 77 (Suppl. 9), 50–65.

Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27 (3), 343–350.

Bernsel, A., Viklund, H., Falk, J., Lindahl, E., von Heijne, G., Elofsson, A., 2008. Prediction of membrane-protein topology from first principles. Proc. Natl. Acad. Sci. U S A 105 (20), 7177–7181.

Beswick, V., Isvoran, A., Nedellec, P., Sanson, A., Jamin, N., 2011. Membrane interface composition drives the structure and the tilt of the single transmembrane helix protein pmp1: Md studies. Biophys. J. 100 (7), 1660–1667.

Bigelow, H., Rost, B., 2006. Proftmb: a web server for predicting bacterial transmembrane beta barrel proteins. Nucleic Acids Res. 34 (Web Server issue), W186–W188.

Bill, R.M., Henderson, P.J., Iwata, S., Kunji, E.R., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G., Vogel, H., 2011. Overcoming barriers to membrane protein structure determination. Nat. Biotechnol. 29 (4), 335–340.

Bond, P.J., Faraldo-Gomez, J.D., Deol, S.S., Sansom, M.S., 2006. Membrane protein dynamics and detergent interactions within a crystal: a simulation study of ompa. Proc. Natl. Acad. Sci. U S A 103 (25), 9518–9523.

Bond, P.J., Holyoake, J., Ivetac, A., Khalid, S., Sansom, M.S., 2007. Coarse-grained molecular dynamics simulations of membrane proteins and peptides. J. Struct. Biol. 157 (3), 593–605.

Bowie, J.U., 2005. Solving the membrane protein folding problem. Nature 438 (7068), 581–589.

Bowler, M.W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., Mueller-Dieckmann, C., Gordon, E.J., Flot, D., McSweeney, S.M., Leonard, G.A., 2010. Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. Acta Crystallogr. D Biol. Crystallogr. 66 (Pt 8), 855–864.

Cheng, J., Baldi, P., 2007. Improved residue contact prediction using support vector machines and a large feature set. BMC Bioinf. 8, 113.

Cheng, Y., Liu, A., Guo, Y., Qi, X., 2004. Computational models for the helix tilt angle. J. Chem. Inf. Comput. Sci. 44 (6), 1952–1960.

Chetwynd, A.P., Scott, K.A., Mokrab, Y., Sansom, M.S.P., 2008. CGDB: a database of membrane protein/lipid interactions by coarse-grained molecular dynamics simulations. Mol. Membr. Biol. 25 (8), 662–669.

Cuff, A.L., Sillitoe, I., Lewis, T., Clegg, A.B., Rentzsch, R., Furnham, N., Pellegrini-Calace, M., Jones, D., Thornton, J., Orengo, C.A., 2011. Extending cath: increasing coverage of the protein structure universe and linking structure with function. Nucleic Acids Res. 39 (Database issue), D420–D426.

Deisenhofer, J., Epp, O., Miki, K., Huber, R., Michel, H., 1985. Structure of the protein subunits in the photosynthetic reaction centre of rhodopseudomonas viridis at 3 Å resolution. Cell Mol. Life Sci. 318, 618624.

Domanski, J., Stansfeld, P.J., Sansom, M.S., Beckstein, O., 2010. Lipidbook: a public repository for force-field parameters used in membrane simulations. J. Membr. Biol. 236 (3), 255–258.

Donizelli, M., Djite, M.A., Novere, N.L., 2006. LGICdb: a manually curated sequence database after the genomes. Nucleic Acids Res. 34 (Database issue), D267–D269.

Dorairaj, S., Allen, T.W., 2007. On the thermodynamic stability of a charged arginine side chain in a transmembrane helix. Proc. Natl. Acad. Sci. U S A 104 (12), 4943–4948.

Dougherty, D.A., 1996. Cation–pi interactions in chemistry and biology: a new view of benzene, phe, tyr, and trp. Science 271 (5246), 163–168.

Doyle, D.A., 2004. Structural themes in ion channels. Eur. Biophys. J. 33 (3), 175–179.

Duque, D., Li, X., Katsov, K., Schick, M., 2002. Molecular theory of hydrophobic mismatch between lipids and peptides. J. Chem. Phys. 116, 10478.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., Sali, A. 2006. Comparative protein structure modeling using modeller. Curr. Protoc. Bioinf. (chapter 5), Unit 5.6.

Fariselli, P., Martelli, P.L., Casadio, R., 2005. A new decoding algorithm for hidden markov models improves the prediction of the topology of all-beta membrane proteins. BMC Bioinf. 6 (Suppl. 4), S12.

Forrest, L.R., Tang, C.L., Honig, B., 2006. On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. Biophys. J. 91 (2), 508–517.

Freddolino, P.L., Kalani, M.Y., Vaidehi, N., Floriano, W.B., Hall, S.E., Trabanino, R.J., Kam, V.W., Goddard, W.A. r., 2004. Predicted 3d structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists. Proc. Natl. Acad. Sci. U S A 101 (9), 2736–2741.

Freeman, T.C.J., Wimley, W.C., 2010. A highly accurate statistical approach for the prediction of transmembrane beta-barrels. Bioinformatics 26 (16), 1965–1974.

Freites, J.A., Tobias, D.J., White, S.H., 2006. A voltage-sensor water pore. Biophys. J. 91 (11), L90–L92.

Friemann, R., Larsson, D.S., Wang, Y., van der Spoel, D., 2009. Molecular dynamics simulations of a membrane protein-micelle complex in vacuo. J. Am. Chem. Soc. 131 (46), 16606–16607.

Fuchs, A., Frishman, D., 2010. Structural comparison and classification of alpha-helical transmembrane domains based on helix interaction patterns. Proteins 78 (12), 2587–2599.

Fuchs, A., Kirschner, A., Frishman, D., 2009. Prediction of helix–helix contacts and interacting helices in polytopic membrane proteins using neural networks. Proteins 74 (4), 857–871.

Gajewski, C., Dagcan, A., Roux, B., Deutsch, C., 2011. Biogenesis of the pore architecture of a voltage-gated potassium channel. Proc. Natl. Acad. Sci. U S A 108 (8), 3240–3245.

Gallin, W.J., Boutet, P.A., 2011. Vkcdb: voltage-gated k+ channel database updated and upgraded. Nucleic Acids Res. 39 (Database issue), D362–D366.

Goldfeld, D.A., Zhu, K., Beuming, T., Friesner, R.A., 2011. Successful prediction of the intra- and extracellular loops of four G-protein-coupled receptors. Proc. Natl. Acad. Sci. U S A. 108 (20), 8275–8280.

Granseth, E., von Heijne, G., Elofsson, A., 2005. A study of the membrane–water interface region of membrane proteins. J. Mol. Biol. 346 (1), 377–385.

Gratkowski, H., Lear, J.D., DeGrado, W.F., 2001. Polar side chains drive the association of model transmembrane peptides. Proc. Natl. Acad. Sci. U S A 98 (3), 880–885.

Gumbart, J., Chipot, C., Schulten, K., 2011. Free-energy cost for translocon-assisted insertion of membrane proteins. Proc. Natl. Acad. Sci. U S A 108 (9), 3596–3601.

Gumbart, J., Schulten, K., 2006. Molecular dynamics studies of the archaeal translocon. Biophys. J. 90 (7), 2356–2367.

Gumbart, J., Wang, Y., Aksimentiev, A., Tajkhorshid, E., Schulten, K., 2005. Molecular dynamics simulations of proteins in lipid bilayers. Curr. Opin. Struct. Biol. 15 (4), 423–431.

Gurezka, R., Laage, R., Brosig, B., Langosch, D., 1999. A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. J. Biol. Chem. 274 (14), 9265–9270.

Hall, S.E., Roberts, K., Vaidehi, N., 2009. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. J. Mol. Graph. Model 27 (8), 944–950.

Harrington, S.E., Ben-Tal, N., 2009. Structural determinants of transmembrane helical proteins. Structure 17 (8), 1092–1103.

Hessa, T., Meindl-Beinker, N.M., Bernsel, A., Kim, H., Sato, Y., Lerch-Bader, M., Nilsson, I., White, S.H., von Heijne, G., 2007. Molecular code for transmembrane-helix recognition by the sec61 translocon. Nature 450 (7172), 1026–1030.

Hildebrand, P.W., Goede, A., Bauer, R.A., Gruening, B., Ismer, J., Michalsky, E., Preissner, R., 2009. SuperLooper – a prediction server for the modeling of loops in globular and membrane proteins. Nucleic Acids Res. 37 (Web Server issue), W571–W574.

Hildebrand, P.W., Gunther, S., Goede, A., Forrest, L., Frommel, C., Preissner, R., 2008. Hydrogen-bonding and packing features of membrane proteins: functional implications. Biophys. J. 94 (6), 1945–1953.

Ho, B.K., Gruswitz, F., 2008. Hollow: generating accurate representations of channel and interior surfaces in molecular structures. BMC Struct. Biol. 8, 49.

Huang, L.S., Sun, G., Cobessi, D., Wang, A.C., Shen, J.T., Tung, E.Y., Anderson, V.E., Berry, E.A., 2006. 3-Nitropropionic acid is a suicide inhibitor of mitochondrial respiration that, upon oxidation by complex ii, forms a covalent adduct with a catalytic base arginine in the active site of the enzyme. J. Biol. Chem. 281 (9), 5965–5972.

Illergard, K., Callegari, S., Elofsson, A., 2010. Mprap: an accessibility predictor for a-helical transmembrane proteins that performs well inside and outside the membrane. BMC Bioinf. 11, 333.

Illergard, K., Kauko, A., Elofsson, A., 2011. Why are polar residues within the membrane core evolutionary conserved? Proteins 79 (1), 79–91.

Ivetac, A., Sansom, M.S., 2008. Molecular dynamics simulations and membrane protein structure quality. Eur. Biophys. J. 37 (4), 403–409.

Izarzugaza, J.M.G., Grana, O., Tress, M.L., Valencia, A., Clarke, N.D., 2007. Assessment of intramolecular contact predictions for CASP7. Proteins 69 (Suppl 8), 152–158.

Jayasinghe, S., Hristova, K., White, S.H., 2001a. Energetics, stability, and prediction of transmembrane helices. J. Mol. Biol. 312 (5), 927–934.

Jayasinghe, S., Hristova, K., White, S.H., 2001b. MPtopo: a database of membrane protein topology. Prot. Sci. 10 (2), 455–458.
Jeon, J., Voth, G.A., 2008. Gating of the mechanosensitive channel protein mscl: the interplay of membrane and protein. Biophys. J. 94 (9), 3497–3511.
Joh, N.H., Min, A., Faham, S., Whitelegge, J.P., Yang, D., Woods, V.L., Bowie, J.U., 2008. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. Nature 453 (7199), 1266–1270.
Johansson, A.C., Lindahl, E., 2006. Amino-acid solvation structure in transmembrane helices from molecular dynamics simulations. Biophys. J. 91 (12), 4450–4463.
Johnson, R.M., Hecht, K., Deber, C.M., 2007. Aromatic and cation–pi interactions enhance helix–helix association in a membrane environment. Biochemistry 46 (32), 9208–9214.
Jones, D.T., 1997. Successful ab initio prediction of the tertiary structure of NK-lysin using multiple sequences and recognized supersecondary structural motifs. Proteins (Suppl. 1), 185–191.
Jones, D.T., 2001. Predicting novel protein folds by using FRAGFOLD. Proteins (Suppl. 5), 127–132.
Jones, D.T., 2007. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. Bioinformatics 23 (5), 538–544.
Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., Krauss, N., 2001. Three-dimensional structure of cyanobacterial photosystem i at 2.5 a resolution. Nature 411 (6840), 909–917.
Käll, L., Krogh, A., Sonnhammer, E.L., 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338 (5), 1027–1036.
Käll, L., Krogh, A., Sonnhammer, E.L., 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. Bioinformatics (Suppl. 1), i251–i257.
Kauko, A., Illergard, K., Elofsson, A., 2008. Coils in the membrane core are conserved and functionally important. J. Mol. Biol. 380 (1), 170–180.
Kelm, S., Shi, J., Deane, C.M., 2010. Medeller: homology-based coordinate generation for membrane proteins. Bioinformatics 26 (22), 2833–2840.
Kim, T., Im, W., 2010. Revisiting hydrophobic mismatch with free energy simulation studies of transmembrane helix tilt and rotation. Biophys. J. 99 (1), 175–183.
Krishna, A.G., Menon, S.T., Terry, T.J., Sakmar, T.P., 2002. Evidence that helix 8 of rhodopsin acts as a membrane-dependent conformational switch. Biochemistry 41 (26), 8298–8309.
Lancaster, C.R., Kroger, A., Auer, M., Michel, H., 1999. Structure of fumarate reductase from wolinella succinogenes at 2.2 a resolution. Nature 402, 377–385.
Langelaan, D.N., Wieczorek, M., Blouin, C., Rainey, J.K., 2010. Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. J. Chem. Inf. Model 50, 2213–2220.
Laskowski, R.A., Macarthur, M.W., Moss, D.S., Thornton, J.M., 1993. Procheck – a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291.
Lee, J.K., Kozono, D., Remis, J., Kitagawa, Y., Agre, P., Stroud, R.M., 2005. Structural basis for conductance by the archaeal aquaporin aqpm at 1.68 a. Proc. Natl. Acad. Sci. U S A 102, 18932–18937.
Lemmon, M.A., Flanagan, J.M., Hunt, J.F., Adair, B.D., Bormann, B.J., Dempsey, C.E., Engelman, D.M., 1992. Glycophorin a dimerization is driven by specific interactions between transmembrane alpha-helices. J. Biol. Chem. 267, 7683–7689.
Li, Y., Goddard, W.A. r., 2008. Prediction of structure of g-protein coupled receptors and of bound ligands with applications for drug design. Pac. Symp. Biocomput., 344–353.
Lindahl, E., Sansom, M.S., 2008. Membrane proteins: molecular dynamics simulations. Curr. Opin. Struct. Biol. 18 (4), 425–431.
Lo, A., Cheng, C.W., Chiu, Y.Y., Sung, T.Y., Hsu, W.L., 2011. Tmpad: an integrated structural database for helix-packing folds in transmembrane proteins. Nucleic Acids Res. 39 (Database issue), D347–D355.
Lo, A., Chiu, Y.Y., Rodland, E.A., Lyu, P.C., Sung, T.Y., Hsu, W.L., 2009. Predicting helix–helix interactions from residue contacts in membrane proteins. Bioinformatics 25 (8), 996–1003.
Lomize, A.L., Pogozheva, I.D., Lomize, M.A., Mosberg, H.I., 2006a. Positioning of proteins in membranes: a computational approach. Prot. Sci. 15 (6), 1318–1333.
Lomize, M.A., Lomize, A.L., Pogozheva, I.D., Mosberg, H.I., 2006b. OPM: orientations of proteins in membranes database. Bioinformatics 22 (5), 623–625.
MacCallum, J.L., Bennett, W.F., Tieleman, D.P., 2008. Distribution of amino acids in a lipid bilayer from computer simulations. Biophys. J. 94 (9), 3393–3404.
Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., de Vries, A.H., 2007. The martini force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B 111 (27), 7812–7824.
Melen, K., Krogh, A., von Heijne, G., 2003. Reliability measures for membrane protein topology prediction algorithms. J. Mol. Biol. 327 (3), 735–744.
Michino, M., Abola, E., Brooks, C.L. r., Dixon, J.S., Moult, J., Stevens, R.C., 2008. Community-wide assessment of gpcr structure modelling and ligand docking: Gpcr dock 2008. Nat. Rev. Drug Discov. 8 (6), 455–463.
Michino, M., Chen, J., Stevens, R.C., Brooks, C.L. r., 2010. Foldgpcr: structure prediction protocol for the transmembrane domain of g protein-coupled receptors from class a. Proteins 78 (10), 2189–2201.
Murata, K., Mitsuoka, K., Hirai, T., Walz, T., Agre, P., Heymann, J.B., Engel, A., Fujiyoshi, Y., 2000. Structural determinants of water permeation through aquaporin-1. Nature 407 (6804), 599–605.
Nanda, V., Schmiedekamp, A., 2008. Are aromatic carbon donor hydrogen bonds linear in proteins. Proteins 70 (2), 489–497.
Neumann, S., Fuchs, A., Mulkidjanian, A., Frishman, D., 2010. Current status of membrane protein structure classification. Proteins 78 (7), 1760–1773.
Norman, K.E., Nymeyer, H., 2006. Indole localization in lipid membranes revealed by molecular simulation. Biophys. J. 91 (6), 2046–2054.
Nugent, T., Jones, D.T., 2009. Transmembrane protein topology prediction using support vector machines. BMC Bioinf. 10, 159.
Nugent, T., Jones, D.T., 2010. Predicting transmembrane helix packing arrangements using residue contacts and a force-directed algorithm. PLoS Comput. Biol. 6 (3), e1000714.
Orgel, J.P., 2004. Sequence context and modified hydrophobic moment plots help identify 'horizontal' surface helices in transmembrane protein structure prediction. J. Struct. Biol. 148 (1), 51–65.
Ou, Y.Y., Chen, S.A., Gromiha, M.M., 2010. Prediction of membrane spanning segments and topology in beta-barrel membrane proteins at better accuracy. J. Comput. Chem. 31 (1), 217–223.
Pellegrini-Calace, M., Carotti, A., Jones, D.T., 2003. Folding in lipid membranes (FILM): a novel method for the prediction of small membrane protein 3D structures. Proteins 50 (4), 537–545.
Pellegrini-Calace, M., Maiwald, T., Thornton, J.M., 2009. Porewalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. PLoS Comput. Biol. 5 (7), e1000440.
Periole, X., Huber, T., Marrink, S.J., Sakmar, T.P., 2007. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. J. Am. Chem. Soc. 129 (33), 10126–10132.
Punta, M., Rost, B., 2005. PROFcon: novel prediction of long-range contacts. Bioinformatics 21 (13), 2960–2968.
Raman, P., Cherezov, V., Caffrey, M., 2006a. The membrane protein data bank. Cell Mol. Life Sci. 63 (1), 36–51.
Raman, P., Cherezov, V., Caffrey, M., 2006b. The Membrane Protein Data Bank. Cell Mol. Life Sci. 63 (1), 36–51.
Randall, A., Cheng, J., Sweredoski, M., Baldi, P., 2008. Tmbpro: secondary structure beta-contact and tertiary structure prediction of transmembrane beta-barrel proteins. Bioinformatics 24 (4), 513–520.
Ray, A., Lindahl, E., Wallner, B., 2010. Model quality assessment for membrane proteins. Bioinformatics 26 (24), 3067–3074.
Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., 2004. Protein structure prediction using Rosetta. Methods Enzymol. 383, 66–93.
Roux, B., 2010. Exploring the ion selectivity properties of a large number of simplified binding site models. Biophys. J. 98 (12), 2877–2885.
Roux, B., Berneche, S., Egwolf, B., Lev, B., Noskov, S.Y., Rowley, C.N., Yu, H., 2011. Perspectives on: ion selectivity in channels and transporters. J. Gen. Physiol. 137 (5), 415–426.
Roux, B., Schulten, K., 2004. Computational studies of membrane channels. Structure 12 (8), 1343–1351.
Saier, M.H.J., Tran, C.V., Barabote, R.D., 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res. 34 (Database issue), D181–D186.
Sal-Man, N., Gerber, D., Bloch, I., Shai, Y., 2007. Specificity in transmembrane helix–helix interactions mediated by aromatic residues. J. Biol. Chem. 282 (27), 19753–19761.
Sands, Z.A., Sansom, M.S., 2007. How does a voltage sensor interact with a lipid bilayer? simulations of a potassium channel domain. Structure 15 (2), 235–244.
Sankararamakrishnan, R., Weinstein, H., 2000. Molecular dynamics simulations predict a tilted orientation for the helical region of dynorphin a(1–17) in dimyristoylphosphatidylcholine bilayers. Biophys. J. 79 (5), 2331–2344.
Sansom, M.S.P., Scott, K.A., Bond, P.J., 2008. Coarse-grained simulation: a high-throughput computational approach to membrane proteins. Biochem. Soc. Trans. 36 (Pt 1), 27–32.
Schneider, D., Engelman, D.M., 2004. Motifs of two small residues can assist but are not sufficient to mediate transmembrane helix interactions. J. Mol. Biol. 343 (4), 799–804.
Schow, E.V., Freites, J.A., Cheng, P., Bernsel, A., von Heijne, G., White, S.H., Tobias, D.J., 2011. Arginine in membranes: the connection between molecular dynamics simulations and translocon-mediated insertion experiments. J. Membr. Biol. 239 (1-2), 35–48.
Scott, K.A., Bond, P.J., Ivetac, A., Chetwynd, A.P., Khalid, S., Sansom, M.S., 2008. Coarse-grained md simulations of membrane protein-bilayer self-assembly. Structure 16 (4), 621–630.
Seeger, M.A., Schiefner, A., Eicher, T., Verrey, F., Diederichs, K., Pos, K.M., 2006. Structural asymmetry of acrb trimer suggests a peristaltic pump mechanism. Science 313 (5791), 295–1298.
Senes, A., Engel, D.E., DeGrado, W.F., 2004. Folding of helical membrane proteins: the role of polar, gxxxg-like and proline motifs. Curr. Opin. Struct. Biol. 14 (4), 465–479.
Shi, Z., Olson, C.A., Kallenbach, N.R., 2002. Cation–pi interaction in model alpha-helical peptides. J. Am. Chem. Soc. 124 (13), 3284–3291.
Simons, K.T., Bonneau, R., Ruczinski, I., Baker, D., 1999. Ab initio protein structure prediction of casp iii targets using rosetta. Proteins Suppl. 3, 171–176.
Simons, K.T., Kooperberg, C., Huang, E., Baker, D., 1997. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. J. Mol. Biol. 268 (1), 209–225.
Singh, N.K., Goodman, A., Walter, P., Helms, V., Hayat, S., 2011. Tmbhmm: A frequency profile based hmm for predicting the topology of transmembrane beta barrel proteins and the exposure status of transmembrane residues. Biochim. Biophys. Acta 1814 (5), 664–670.

Stewart, J.J., 2009. Application of the pm6 method to modeling proteins. J. Mol. Model 15 (7), 765–805.
Treptow, W., Marrink, S.J., Tarek, M., 2008. Gating motions in voltage-gated potassium channels revealed by coarse-grained molecular dynamics simulations. J. Phys. Chem. B 112 (11), 3277–3282.
Tusnady, G.E., Dosztanyi, Z., Simon, I., 2005a. PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res. 33 (Database issue), D275–D278.
Tusnady, G.E., Dosztanyi, Z., Simon, I., 2005b. TMDET: web server for detecting transmembrane regions of proteins by using their 3D coordinates. Bioinformatics 21 (7), 1276–1277.
Tusnady, G.E., Kalmar, L., Simon, I., 2008. TOPDB: topology data bank of transmembrane proteins. Nucleic Acids Res. 36 (Database issue), D234–D239.
Tusnady, G.E., Simon, I., 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics 17 (9), 849–850.
Viklund, H., Bernsel, A., Skwark, M., Elofsson, A., 2008. SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics 24 (24), 2928–2929.
Viklund, H., Elofsson, A., 2008. OCTOPUS: improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics 24 (15), 662–1668.
Viklund, H., Granseth, E., Elofsson, A., 2006. Structural classification and prediction of reentrant regions in alpha-helical transmembrane proteins: application to complete genomes. J. Mol. Biol. 361 (3), 591–603.
von Heijne, G., 1992. Membrane protein structure prediction Hydrophobicity analysis and the positive-inside rule. J. Mol. Biol. 225 (2), 487–494.
Vostrikov, V.V., Daily, A.E., Greathouse, D.V., Koeppe, R.E., 2010. Charged or aromatic anchor residue dependence of transmembrane peptide tilt. J. Biol. Chem. 285 (41), 31723–31730.
Vriend, G., 1990. What if: a molecular modeling and drug design program. J. Mol. Graph. 8 (1), 52–56, 29.
Vroling, B., Sanders, M., Baakman, C., Borrmann, A., Verhoeven, S., Klomp, J., Oliveira, L., de Vlieg, J., Vriend, G., 2011. Gpcrdb: information system for g protein-coupled receptors. Nucleic Acids Res. 39 (Database issue), D309–D319.
Wallner, B., Elofsson, A., 2006. Identification of correct regions in protein models using structural, alignment, and consensus information. Prot. Sci. 15 (4), 900–913.
Walters, R.F.S., DeGrado, W.F., 2006. Helix-packing motifs in membrane proteins. Proc. Natl. Acad. Sci. U S A 103 (37), 13658–13663.
Wee, C.L., Chetwynd, A., Sansom, M.S., 2011. Membrane insertion of a voltage sensor helix. Biophys. J. 100 (2), 410–419.
White, S.H., 2009. Biophysical dissection of membrane proteins. Nature 459 (7245), 344–346.
White, S.H., Wimley, W.C., 1999. Membrane protein folding and stability: physical principles. Annu. Rev. Biophys. Biomol. Struct. 28, 319–365.
Yamashita, A., Singh, S.K., Kawate, T., Jin, Y., Gouaux, E., 2005. Crystal structure of a bacterial homologue of na+/cl – dependent neurotransmitter transporters. Nature 437 (7056), 215–223.
Yarov-Yarovoy, V., Schonbrun, J., Baker, D., 2006. Multipass membrane protein structure prediction using Rosetta. Proteins 62 (4), 1010–1025.
Yefimov, S., van der Giessen, E., Onck, P.R., Marrink, S.J., 2008. Mechanosensitive membrane channels in action. Biophys. J. 94 (8), 2994–3002.
Yohannan, S., Faham, S., Yang, D., Whitelegge, J.P., Bowie, J.U., 2004. The evolution of transmembrane helix kinks and the structural diversity of g protein-coupled receptors. Proc. Natl. Acad. Sci. U S A 101 (4), 959–963.
Yu, H., Noskov, S.Y., Roux, B., 2010. Two mechanisms of ion selectivity in protein binding sites. Proc. Natl. Acad. Sci. U S A 107 (47), 20329–20334.
Zhang, Y., Devries, M.E., Skolnick, J., 2006. Structure modeling of all identified g protein-coupled receptors in the human genome. PLoS Comput. Biol. 2 (2), e13.
Zhou, F.X., Merianos, H.J., Brunger, A.T., Engelman, D.M., 2001. Polar residues drive association of polyleucine transmembrane helices. Proc. Natl. Acad. Sci. U S A 98 (5), 2250–2255.