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.

© 2011 Elsevier Inc. All rights reserved.

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

Download English Version:

https://daneshyari.com/en/article/5914692

Download Persian Version:

https://daneshyari.com/article/5914692

Daneshyari.com