Biochimica et Biophysica Acta 1859 (2017) 2021–2039

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Review

Membrane proteins structures: A review on computational modeling tools

Jose G. Almeida [a,1], Antonio J. Preto [a,1], Panagiotis I. Koukos [b], Alexandre M.J.J. Bonvin [b], Irina S. Moreira [a,b,*]

[a] CNC - Center for Neuroscience and Cell Biology, Rua Larga, FMUC, Polo I, 1°andar, Universidade de Coimbra, 3004-517, Coimbra, Portugal
[b] Bijvoet Center for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Utrecht, Padualaan 8, 3584CH, The Netherlands

ARTICLE INFO

Article history:
Received 23 March 2017
Received in revised form 4 July 2017
Accepted 13 July 2017
Available online 15 July 2017

Keywords:
Membrane proteins
Computational modeling
Machine-learning
Transporters
GPCRs

ABSTRACT

Background: Membrane proteins (MPs) play diverse and important functions in living organisms. They constitute 20% to 30% of the known bacterial, archaean and eukaryotic organisms' genomes. In humans, their importance is emphasized as they represent 50% of all known drug targets. Nevertheless, experimental determination of their three-dimensional (3D) structure has proven to be both time consuming and rather expensive, which has led to the development of computational algorithms to complement the available experimental methods and provide valuable insights.
Scope of review: This review highlights the importance of membrane proteins and how computational methods are capable of overcoming challenges associated with their experimental characterization. It covers various MP structural aspects, such as lipid interactions, allostery, and structure prediction, based on methods such as Molecular Dynamics (MD) and Machine-Learning (ML).
Major conclusions: Recent developments in algorithms, tools and hybrid approaches, together with the increase in both computational resources and the amount of available data have resulted in increasingly powerful and trustworthy approaches to model MPs.
General significance: Even though MPs are elementary and important in nature, the determination of their 3D structure has proven to be a challenging endeavor. Computational methods provide a reliable alternative to experimental methods. In this review, we focus on computational techniques to determine the 3D structure of MP and characterize their binding interfaces. We also summarize the most relevant databases and software programs available for the study of MPs.

Contents

* Corresponding author at: CNC - Center for Neuroscience and Cell Biology, Rua Larga, FMUC, Polo I, 1°andar, Universidade de Coimbra, 3004-517, Coimbra, Portugal.
E-mail address: irina.moreira@cnc.uc.pt (I.S. Moreira).
[1] Equal contribution.

http://dx.doi.org/10.1016/j.bbamem.2017.07.008
0005-2736/

1. Introduction

Membrane proteins (MPs) have diverse functional roles, featuring important functions such as ion and molecule transport, immune system molecule recognition and energy transduction [1]. It is therefore fundamental to comprehensively understand their structure and structure-function relationships. 3D structures of various MPs have been characterized in recent years by several experimental methods, such as Nuclear Magnetic Resonance (NMR), X-ray crystallography and cryo-electron microscopy [2]. MPs, unlike soluble proteins, are difficult to analyze in their native environment, due to their insertion in the lipidic membrane [2–3]. They are affected by the membrane and various specific factors, such as cholesterol content [4] and hydrophobic thickness of the lipid bilayer [5], but also influence the membrane structure itself [5a]. All these aspects contribute to the technical experimental difficulties in the structural characterization of MPs, which explains their relatively low number in the Protein Data Bank (PDB) [6], despite their high proportion in the human proteome [7].

The computational prediction of soluble protein structure can be considered a particularly advanced field, both in terms of variety of approaches and the accuracy they can achieve [8]. However, computational prediction of MPs and their interfaces, especially when studying dimers or high-order oligomers, is still in its early days [9]. Current approaches are usually based on a combination of homology modeling [10] or de novo protein structure determination [11] with ML algorithms [12] to predict binding interfaces and/or intermolecular contacts, and MD simulations to refine the models and study their dynamical properties [13].

Some MPs are of particular interest for therapy assessment and drug targeting given their role in physiological processes and biochemical pathways. Among them are G-protein Coupled Receptors (GPCRs), ion channels and transporters. All these cover a wide array of functions while maintaining some common traits among their respective (super)families. Here, we aim at giving a brief overview of MP and the experimental methods for determining their structure, followed by a comprehensive assessment of known computational methods for the prediction of MP structure and structure-related characteristics, such as topology and binding interface prediction. Lastly, we highlight some recent computational studies on key MPs and their main features.

2. Membrane proteins

MPs have been defined as proteins associated to lipid domains, which are involved in communication, regulation and structural coherence. In fact, proteins that entirely or partially span the membrane (intrinsic/Trans membrane (TM) proteins), as well as proteins that are peripherally membrane-bound (peripheral MPs – PMPs), can carry out these functions. Due to the high amount of information and computational methods for MPs, we focused on TM proteins, which will be referred to as MPs. For readers interested in PMPs, specialized reviews can be found covering this class of membrane proteins [14], their interaction with the membrane [15] and the experimental and computational methods for their study [16].

Only a detailed understanding of MP structure-function relationships will allow the understanding of common pathologies at a molecular level and the development of improved pharmacological procedures [17,18,19]. The most functionally relevant intrinsic MPs are typically split into ion channels, membrane receptors and transporters [1a,20]. Ion channels facilitate the diffusion of ions across membranes, bridging the intra- and extracellular environments across the hydrophobic lipid bilayer by allowing hydrophilic molecules and ions to pass through the membrane. Ion channels are structurally modulated by the TM electrochemical potential, the binding of ligands, and mechanical stress and/or changes in the local lipid environment [21]. In some cases, this modulation is required for biological function [22]. Membrane receptors, comprising GPCRs as well as olfactory receptors (ORs) and nuclear receptors [23], play roles in different biochemical and signaling pathways, and in triggering environment, immune, hormonal and neurological responses, which makes them highly interesting targets for therapeutical investigation. They often share common structural traits, allowing for their classification into protein families or superfamilies. Transporters span the cell membrane with recurring specific membrane topologies, energy coupling mechanisms and substrate specificities. They are capable of transporting molecules and ions across the membrane, triggering environment-driven responses, delivering essential nutrients and disposing cellular waste.

MPs as defined in this review consist typically of several domains: extracellular (typically involved in cell-cell signaling and/or interactions), intracellular (performing a wide range of functions such as activating signaling pathways and anchoring cytoskeletal proteins) and intramembrane (such as pores and channels) [24]. TM proteins in general are amphipathic, meaning that they have different electronegativity and hydrophobicity profiles along their structure, allowing them to be both in contact with water (hydrophilic environment) and the membrane (hydrophobic environment). The structure and function of many TM proteins depend on Post Translational Modifications (PTM) such as phosphorylation and glycosylation. The two major recurrent protein structure motifs in MPs are TM α-helices [25], repeatedly crossing the membranes in α-helical bundles and β-strands arranged into super-secondary structures known as β-barrels [26].

3. Experimental structural determination of membrane proteins

Despite their functional importance, only 4.193 structures of membrane proteins (or rather of sub-domains) can be found among the 131.485 determined protein structures deposited at the PDB [7] (statistics from June 29th 2017) (Fig. 1). This means that <1% of all determined protein structures belong to MP families. This number includes multiple submissions of the same protein under a variety of experimental conditions. In contrast to the limited number of available MP 3D structures, there are 199.322 MP sequence clusters according to UniProt's UniRef (June 29th 2017).

Two major factors can explain this discrepancy: i) difficulties in both expression, which can be done in several organisms [27] but mostly in *Escherichia coli* (*E. coli*) [28] and purification processes; ii) challenges associated with the actual determination of the 3D structure of the