Biochimica et Biophysica Acta 1858 (2016) 2468–2482

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

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

Efficient preparation and analysis of membrane and membrane protein systems☆

Matti Javanainen [a,*,1], Hector Martinez-Seara [a,b,*,1]

[a] Department of Physics, Tampere University of Technology, Tampere, Finland
[b] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Prague, Czech Republic

ARTICLE INFO

Article history:
Received 30 December 2015
Received in revised form 23 February 2016
Accepted 25 February 2016
Available online 4 March 2016

Keywords:
Tools and software
Membrane building
Protein insertion
Molecular dynamics
Lipid bilayer

ABSTRACT

Molecular dynamics (MD) simulations have become a highly important technique to consider lipid membrane systems, and quite often they provide considerable added value to laboratory experiments. Rapid development of both software and hardware has enabled the increase of time and size scales reachable by MD simulations to match those attainable by several accurate experimental techniques. However, until recently, the quality and maturity of software tools available for building membrane models for simulations as well as analyzing the results of these simulations have seriously lagged behind.

Here, we discuss the recent developments of such tools from the end-users' point of view. In particular, we review the software that can be employed to build lipid bilayers and other related structures with or without embedded membrane proteins to be employed in MD simulations. Additionally, we provide a brief critical insight into force fields and MD packages commonly used for membrane and membrane protein simulations. Finally, we list analysis tools that can be used to study the properties of membrane and membrane protein systems. In all these points we comment on the respective compatibility of the covered tools.

We also share our opinion on the current state of the available software. We briefly discuss the most commonly employed tools and platforms on which new software can be built. We conclude the review by providing a few ideas and guidelines on how the development of tools can be further boosted to catch up with the rapid pace at which the field of membrane simulation progresses. This includes improving the compatibility between software tools and promoting the openness of the codes on which these applications rely.

This article is part of a Special Issue entitled: Biosimulations edited by Ilpo Vattulainen and Tomasz Róg.

1. Introduction

The first simulation on soft matter was performed 30 years ago, and since then the field of computational biophysics has expanded at an enormous pace. This first Monte Carlo simulation, studying the water–lipid interface [1], was followed by studies on micelles [2] and bilayers [3,4] using the molecular dynamics (MD) method. Simulations of membrane proteins took place soon after [5,6].

Since these ground-breaking studies in the early and mid 90s, both computing power and the accuracy of the employed models have increased drastically, leading to a large number of studies on membranes (see *e.g.* [7,8]) and membrane protein systems (see *e.g.* [9–11]). What is more, experimental techniques have also improved, providing more accurate data against which the simulation models can be parameterized and optimized. Nowadays the knowledge required to perform MD simulations of membranes or membrane protein systems is easily available for everyone *via* the internet. Such simulations can be performed with numerous available software packages, including several free reliable alternatives, on any modern desktop computer to a certain extent.

However, except for the last few years, what has been seriously lacking are publicly available user-friendly tools that aid the setting up and analysis of membrane or membrane protein simulations. Such tools are necessary to make the field of computational biophysics more approachable to newcomers. Additionally, they would also simplify the tasks of experienced scientists, as automation and ease-of-use of tools would leave more time for the actual science. Luckily things are changing and a number of new approaches have been introduced to both building lipid membranes and inserting proteins into them, as well as to analyzing the results of the simulations on these systems.

Most of these new tools have been made available since the last thorough review on the topic almost ten years ago [12], which calls for an update. In this paper we review the important aspects of setting up and analyzing membrane and membrane protein simulations. It should be noted that this review does not aim to provide step-by-step instructions for performing membrane or membrane protein simulations, yet such recipes are available in *e.g.* Refs. [13] and [14]. Instead,

☆ This article is part of a Special Issue entitled: Biosimulations edited by Ilpo Vattulainen and Tomasz Róg.

* Corresponding authors.
E-mail addresses: matti.javanainen@tut.fi (M. Javanainen), hseara@gmail.com (H. Martinez-Seara).

[1] Both authors contributed equally to the article.

http://dx.doi.org/10.1016/j.bbamem.2016.02.036
0005-2736/

we aim to provide a comprehensive list of the key software available. We comment the ease-of-use and generality of these tools and also provide information on their compatibility with force fields and file formats. This listing will aid both newcomers to select the proper tools for their project as well as inform more experienced users of newly published tools and techniques. It must be noted that the ever increasing user-friendliness of the applications and simulation software might, however, introduce a new and a perhaps surprising issue. Newcomers without the proper background knowledge on the underlying algorithms might nowadays be able to perform both simulations and analyses. This might accidentally lead to incorrect conclusions that are extremely hard to catch during the peer-review process. Therefore, it is important that regardless of how easy to use scientific tools become, they should never be used as black boxes.

This review is structured as follows. We first introduce the most common force fields employed in molecular dynamics simulations of lipids and proteins. Next, the numerous approaches used to build lipid bilayers are reviewed. This is followed by a thorough list of techniques and tools for the insertion of proteins into membranes. After a brief examination of the popular molecular dynamics software packages, we review tools available for the analysis of membrane and membrane protein simulations. Finally, we raise issues related to the current paradigm of tool development and try to foresee how these issues could be tackled in the near future.

2. Force fields for biomolecular simulations

A careful selection of the proper lipid and protein force fields is of key importance for every project considering MD simulations on biomolecular systems. Most importantly, the level of detail of the chosen force field, be it *e.g.* a fully atomistic or a coarse-grained one, should allow to sample time and size scales relevant for the problem at hand yet still provide the required chemical accuracy. Another factor affecting the selection of the force field is its compatibility with the available simulation software. What is more, the chosen force fields should either include the molecule parameters related to the research problem or provide tools for parameterizing them. Lipid force fields seldom cover all possible head groups and tail types. Notably, certain head groups (such as phosphatidylcholine) and tails (such as palmitic acid or oleic acid) are often parameterized first and appear in almost every lipid force field. On the other hand, some head groups (such as phosphatidylinositol) or tails (such as linoleyl or linolenoyl) are rarely available. Therefore, the desired membrane composition might limit the number of plausible force fields. The choice of the lipid force field also often sets limits to the available options for the protein force field, and *vice versa*. Sometimes the projects involve molecules beyond lipid and protein families (such as sugars or nucleotides) and in such cases the selected force field should also cover these extra molecule types or be compatible with a force field that contains them.

Some common force field models, which can be divided into different categories based on how much detail they provide, are briefly listed below. For more thorough reviews and comparisons of lipid and protein force fields please see Refs. [15–19]. Notably, no thorough comparison of the performance of the force fields in describing membrane protein systems exists in the literature to our knowledge.

2.1. Coarse-grained models

Coarse-grained models map multiple atoms into larger pseudoatoms or "beads", which significantly reduces the number of degrees of freedom and therefore allows longer simulation times.

The Martini model has gained broad acceptance in the biomolecular simulation community. It contains parameters for lipids [20], including glycolipids [21], and proteins [22,23] as well as carbohydrates [24] and nucleic acids [25] among others. It is also compatible with a polarizable water model [26]. The implicit solvent version of the Martini lipid force field, titled Dry Martini, is also available [27]. One major advantage of Martini, in addition to the large selection of parameterized molecule types, is the number and quality of tools provided on the Martini website.

The PLUM force field also relies on a solvent-free approach and contains parameters for both proteins and lipids [28–30]. One key advantage that PLUM has over Martini is that it describes protein folding, whereas secondary structures are fixed in Martini.

Furthermore, the ELBA force field [31] introduces dipoles into both lipid molecules and water beads, which greatly improves the description of electrostatics. However, the number of lipid types available is very limited and proteins have not been parameterized at the time of writing this review.

2.2. United-atom force fields

United atom models usually combine methyl groups and methylene bridges into pseudoatoms, thus effectively combining the properties of the hydrogen atoms into their host carbons. The most common of such force fields, namely GROMOS, contains multiple parameter sets for proteins with the newest one being 54A7/54B7 [32]. The multiple versions are also compatible with the corresponding lipid force fields [33–35] and contain parameters for many other molecule types, such as carbohydrates and nucleic acids. Two automated web-based tools exist for the parameterization of small molecules for GROMOS. The long-running and popular PRODRG server [36,37] has recently received criticism, most importantly for its poor handling of charge groups [38]. The more recent Automated Topology Builder (ATB) [39,40] aims to tackle the charge group partitioning issue [41], in addition to other improvements.

Here, the commonly employed yet old Berger united atom lipid model [42] should be mentioned. It combines parameters from multiple sources and has been used together with atomistic protein force fields (see below). This parameterization was recently refined to correctly describe phase behavior [43].

The united atom TraPPE force field contains parameters for lipids [44] yet parameters for proteins are not available.

2.3. Atomistic force fields

Thanks to the rise in computing power, researchers can now waive the performance provided by united atom approaches in favor to the improved accuracy provided by fully atomistic force fields. Additionally, the interest towards membrane protein simulations has called for the development of high quality lipid force fields compatible with the protein force fields previously employed in simulations of water-soluble proteins.

Various versions of the Amber protein force field are commonly used, with ff99SB-ILDN [45] gaining widespread acceptance. Additionally, the ff99SB force field was recently refined in the form of ff14SB [46]. Further, another development branch entitled ff14ipq employed charges derived in a new way [47] and has not yet been thoroughly tested. Even the old ff03 is still used to some extend [48] (note that ff03 is from 2003 whereas ff99SB-ILDN is from 2010).

Multiple Amber-compatible sets of lipid parameters also exist. The General Amber Force Field (GAFF) lipid parameters [49] were later combined with the development of Lipid11 [50] resulting in the Lipid14 parameter set [51,52]. Lipid14 contains parameters for several lipid types as well as cholesterol. Until Lipid14 all these force fields required the use of applied surface tension in order to maintain the membrane in a liquid phase. In addition, the Slipids parameter set [53–55] is compatible with the Amber protein force fields and has parameters for multiple lipid types including sphingomyelin and cholesterol. However, polyunsaturated tails are not included in Slipids. Amber also supports the Glycam carbohydrate force field [56]. Automated ways to parameterize molecules, such as drugs, for the GAFF [57] force