MERMAID: dedicated web server to prepare and run coarse-grained membrane protein dynamics

Mangesh Damre1,2,†, Alessandro Marchetto2,† and Alejandro Giorgetti2,3,*

1Neurobiology, International School for Advanced Studies (SISSA), via Bonomea 265, 34136 Trieste, Italy,
2Department of Biotechnology, University of Verona, Ca Vignal 1, strada Le Grazie, 15, 37134 Verona, Italy and
3Institute for Advanced Simulations (IAS)-5/Institute for Neuroscience and Medicine (INM)-9, Forschungszentrum Jülich, 52428 Jülich, Germany

ABSTRACT

Atomistic molecular dynamics simulations of membrane proteins have been shown to be extremely useful for characterizing the molecular features underlying their function, but require high computational power, limiting the understanding of complex events in membrane proteins, e.g. ion channels gating, GPCRs activation. To overcome this issue, it has been shown that coarse-grained approaches, although requiring less computational power, are still capable of correctly describing molecular events underlying big conformational changes in biological systems. Here, we present the Martini coarse-grained membrane protein dynamics (MERMAID), a publicly available web interface that allows the user to prepare and run coarse-grained molecular dynamics (CGMD) simulations and to analyse the trajectories.

INTRODUCTION

A lipid membrane isolates the cells from the external world. It has also a fundamental role in cellular communication because it is the target of all the extracellular stimuli acting on the cell. Several different families of proteins, i.e. integral membrane proteins, are specialized in detecting these signals and in internalizing this information to the cell. The original human genome sequencing project estimated 20% of the total gene count of 31 778 genes to code for membrane proteins (1). This big number shows the importance of membrane proteins for the different organisms. Unfortunately, and principally due to the difficulties at the experimental level, just a small percentage of them have been completely characterized. These difficulties can be overcome by an extensive combination of experimental and in silico studies. In this regard, atomistic molecular dynamics (MD) simulations have been extensively used to study membrane proteins at different time and space scales (2,3). Although the recent advancements in the development of force fields, algorithms and hardware, running long MD simulations to explore large conformational changes is still extremely expensive in terms of time and computational power. Indeed, exploring events that occur in the micro-milliseconds time-scale, like ion channels gating or receptor activation is prohibitive. Different techniques have been developed and implemented over the years in order to accelerate MD simulation studies (4). Among them, coarse-grained molecular dynamics (CGMD) simulations approaches have evolved very rapidly during the last few years and reached important accuracy levels (5). CGMD is computationally less expensive and can explore longer time-scales, still being able to accurately explain molecular events. Indeed, it ultimately results in reducing the degrees of freedom and increase the speed of the simulations. Thus CGMD helps to overcome the experimental limitations to study large scale biomolecular processes over longer time scales.

Due to the advancement in the simulation techniques, in recent years, there is a significant increase in the number of coarse-grained studies (6). Indeed, CG force fields like MARTINI (7), SIRAH (8), Shinoda-Devane-Klein (SDK) (9) and ELBFFs (10), can be run routinely, within several MD simulations packages. Here, we present our Martini coarse-grained membrane protein dynamics, MERMAID web server, aimed at preparing and running CGMD simulations directly from a publicly available web interface (Figure 1). This interface is directed to expert and non-expert users as it allows the preparation of files and allows a heavy interaction with the users. Finally, the user can choose either to run and analyse CGMD simulations directly through our servers or to download the files to be run in local machines.

MATERIALS AND METHODS

Implementation

MERMAID workflow can be divided in two different stages that include the user interface and the backend of the server. In detail:

†To whom correspondence should be addressed. Tel: +39 45 8027905; Fax: +39 45 8027929; Email: alejandro.giorgetti@univr.it

*The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2019. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Figure 1. MERMAID web server interface (Start page, Parameter page): MERMAID provides two ways to submit a protein structure: a) from OPM database, or b) the user can upload the protein structure from a local system. After uploading the protein structure, MERMAID will redirect the user to the parameter page.

Web interface. The user can interact with the client web interface to submit a membrane protein structure (Figure 1). It can be either a custom PDB structure or a structure directly obtained from the Protein Data Bank (PDB) (http://www.rcsb.org) (11). Naturally, most of the PDB structures are not aligned to the z-axis thus, in order to align these protein structures to the membrane normal axis, the user can use the PPM server (https://opm.phar.umich.edu/ppm_server). Moreover, the user is suggested to input a protein structure from the database ‘Orientations of Proteins in Membranes (OPM)’ (https://opm.phar.umich.edu) (12). As the protein secondary structure information is needed to run MARTINI CGMD simulations, MERMAID will calculate it automatically, by using a locally installed version of the Dictionary of Protein Secondary Structure (DSSP) program (13). The user is then funneled to a series of online forms that help in the creation of the parameters files. The user can choose either to keep the default parameters unmodified or to modify them. We implemented our protocol in the GROMACS package with the use of MARTINI force fields. The processes will then run on our server for successful completion of the simulation. MARTINI force fields implemented in this protocol are freely available from http://cgmartini.nl/.

Backend of the server. All the simulations and calculations are performed in the backend of the server (Figure 2). In particular it uses: (i) a local version of Gromacs 4.6 (14) to perform simulations; (ii) a locally installed version of DSSP (13); (iii) all the programs needed for preparing the files for Martini, including martiniize.py (15) and insane.py (16); (iv) GNUPLOT (http://gnuplot.info/) for data plotting and (v) NGL viewer (17) to provide an interactive molecular viewer on the web page.

Protocol

After submission of all the files and parameters, the web server creates an user directory, where all operations will be performed. During the initial setup of the system, the atomistic protein structure is converted into a coarse-grained representation with the help of the martiniize.py python script (15). During this process, all the ligands, ions and atoms not belonging to the protein are removed. MERMAID performs simulations in absence of ligands due to unavailability of appropriate martini forcefield parameters. By default, disulfide bonds present in the proteins are considered during the system preparation. The CG structure is then embedded into a user-defined coarse-grained lipid membrane with the help of the python script insane.py (16). Furthermore, the MERMAID server allows the user to select different kind of lipid membranes of known composition. Some of the offered membranes include: (a) plasma membrane (Human erythrocytes) with a composition of POPC:20, POPE:18, CHOL:20, POPS:7, POPI:3, POSM:18, CDL0:0 and DPG1:3; (b) myelin membrane (Human neurons) with POPC:11, POPE:17, CHOL:28, POPS:9, POPI:1, POI:8, CDL0:0 and DPG1:20; (c) nuclear membrane with POPC:44, POPE:17, CHOL:10, POPS:4, POPI:6, POSM:3, CDL0:1 and DPG1:1; (d) endoplasmic reticulum membrane (Rat) with POPC:48, POPE:19, CHOL:6, POPS:4, POPI:8, POSM:5, CDL0:0 and DPG1:1 and (e) Golgi membrane (rat) with POPC:25, POPE:9, CHOL:8, POPS:3, POPI:5, POSM:7, CDL0:0 and DPG1:0 (18).

Moreover, the user can model a customized symmetric or asymmetric membrane varying the concentration and type of lipids. Normal or polarized water molecules are added during system preparation. Counterions can also be added at this stage of the system preparation. The user has access...
Figure 2. Workflow of the web server. The submitted protein structure is then converted into coarse-grained structure and inserted into CG membrane to form a complex structure. The structure then undergoes minimization, equilibration and production run respectively. Every phase of the coarse-grained simulation is analysed and can be visualised. Lastly, all the results can be downloaded from our server.

to all the generated input and output files for each CGMD process at any time.

As the simulations may take several minutes to be completed, while the simulation is running, it will show a progress bar and the page will refresh continuously until the job is completed. Otherwise, it will show an error message. Thus, the user is constantly updated on the status of the simulation.

The data can be accessed either from the bookmarked web link or directly from the web server user directory by providing the required credentials in the MERMAID search bar. A typical CGMD run within our server consists of four phases:

(i) Minimization run;
(ii) Equilibration run in two different ensembles namely constant NVT and constant NPT;
(iii) Production run continued in constant NPT ensemble;
(iv) Analysis of all the trajectories produced during simulation.

Server architecture

The MERMAID server is equipped with the web interface written in HTML 4/5 (https://www.w3.org/html/), PHP (http://www.php.net/), JavaScript (https://www.javascript.com/), Bash shell (https://www.gnu.org/software/bash/), jQuery (https://jquery.com/), and Bootstrap 4 (https://getbootstrap.com/). For the submitted jobs, the web server prepares MERMAID input files on our local cluster. At last, the results are post-processed (which includes analysis of CGMD trajectory, construction of various data plots) and displayed graphically by using Highcharts (https://www.highcharts.com/) and Plotly (https://plot.ly). The web server can be accessed from any media device and the jobs can be monitored and followed up.

RESULTS

The server has been intensively used by the members of our group and the most representative cases are available as examples to the users. The examples can be retrieved and analysed from Q&A page of MERMAID, through the ‘examples’ section (http://molsim.sci.univr.it/mangesh/questions.html). They include different multimeric states of membrane proteins, i.e. (i) a monomer of the Chloride pumping rhodopsin (PDB accession code: 5B2N); (ii) a dimer of the translocator protein 18 kDa (TSPO) (PDB accession code: 4UC1); (iii) a trimer of the Escherichia coli OmpF porin (PDB accession code: 3POX) and (iv) the tetrameric conformation of the crystal structure of Aquaporin AqpM (PDB accession code: 2F2B). These simulations were generated automatically by just retrieving the structures and running the simulations with default parameters. Here, we describe more in detail case iii), the CGMD simulations of the trimeric structure of the E. coli OmpF porin. In this case, the structure was retrieved from OPM database and funneled through our pipeline (Figure 2).
Figure 3. MERMAID server output. The upper panel shows the trimeric protein structure of the E. coli OmpF porin (PDB accession code: 3POX) in (A) atomistic description, (B) initial coarse-grained structure embedded in a coarse-grained membrane, (C) coarse-grained structure from the last frame of production run. A list of the analysis performed on the CG simulation (D). (E–H) shows the plots for temperature, pressure, system density, and system volume respectively.

Table 1. Analysis performed on CGMD trajectories

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Kinetic energy</th>
<th>Potential energy</th>
<th>Total energy</th>
<th>Temperature</th>
<th>Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enthalpy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PV analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean square displacement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radial distribution function</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We run the protocol (see section Protocol A) of the protein embedded in a human plasma membrane with a composition of POPC:20, POPE:18, CHOL:20, POPS:7, POPL:3, POSM:18, CDL:0:0 and DPG1:3 lipids respectively. The production run was of 100 ns. The system reached the temperature and pressure of equilibrium within the first 10 ns of simulation and maintained the principal structural features along the entire simulation (Figure 3E–H) (details of this simulation can be appreciated here: http://molsim.sci.univr.it/mangesh/RESULT.php?Email=trimer&jobid=20190309-CG_User-0222).

Details of the analysis at the different stages of the simulation are enlisted in Table 1. The initially generated input data and output data can be retrieved from the server. To the best of our knowledge, it is the first server for generating input files as well as performing CGMD using Martini force fields through an online service. The MERMAID web server is freely accessible at http://molsim.sci.univr.it/mermaid/ and an extensive tutorial is available at http://molsim.sci.univr.it/mermaid/tutorial.php.

LIMITATIONS

Some coarse-grained simulation jobs may fail due to missing atoms in the residues of a protein. To overcome this issue, it is recommended to the users to check the protein structures for missing atoms with the help of third-party softwares or with the use of our in-house web server (PRUDE: http://molsim.sci.univr.it/mermaid/prude.php) to add the missing atoms. After addition of missing atoms in the protein structure, the user can upload the complete protein structure to MERMAID. The other limitation resides in orienting the protein to the z-axis of the membrane. The users are strongly suggested to download the protein already aligned from the OPM server (https://opm.phar.umich.edu/) or, if not present, to align it using the PPM server (https://opm.phar.umich.edu/ppm-server). Among the results files, we offer the necessary scripts and programs to prepare the atomistic structure using backmapping resources (http://cgmartini.nl/index.php/back).
CONCLUSION

Computational techniques such as atomistic and CGMD simulations are key to gain deeper insights into the mechanisms underlying the functioning of membrane proteins. Due to the increasing interest and to the availability of more powerful algorithms and hardware, during the last years, several web servers have been developed with the scope of preparing and/or performing different resolution levels of molecular dynamics simulations (Table 2). In particular, the existing web servers (Table 2) provide a web-based graphical user interface to generate various molecular simulation systems and input files to facilitate the usage of common simulation techniques. They are widely used to prepare the input files needed to be run on local clusters, and in some of them to run short simulations. In general, they are not specialized in membrane proteins and they do not allow neither the selection of different membranes nor to run CGMD simulations. Both features, at our advice, are fundamental to get a quick initial idea on the stability of the membrane system and to facilitate the generation of the different steps and protocols of the simulation. Our MERMAID web server implements several novel elements in the ambit of CG simulations, such as the possibility of choosing different membrane composition, but also allows the user to run the initial part of the simulation protocol. The user is free to change the parameters and to choose the different protocols for running simulations. Then the users can choose to run the simulations or to download all the input files to be run locally. If the simulations are run through our servers, a principal set of analysis are carried out in order to follow the evolution of fundamental features, such as temperature, pressure and different energy terms during the simulation (Table 1). Summarizing, the MERMAID server is a user-friendly web server, available to all the users without any login requirement. MERMAID is the only online protocol specifically designed for preparing and running CGMD simulations of membrane proteins through a web server.

FUNDING

Department of Biotechnology, Università degli Studi di Verona (to A.G.). Funding for open access charge: Department of Biotechnology, Università degli Studi di Verona (to A.G.).

Conflict of interest statement. None declared.

REFERENCES


DATA AVAILABILITY

MERMAID is freely accessible at http://molsim.sci.univr.it/mermaid/.

ACKNOWLEDGEMENTS

We deeply acknowledge the contribution of ‘Centro Piattaforme Tecnologiche dell’ Università degli Studi di Verona’.

Table 2. Known web servers to perform MD simulations in all-atom (AA) form and/or in coarse-grained (CG) form in presence or absence of membrane

<table>
<thead>
<tr>
<th>Name</th>
<th>Input Preparation</th>
<th>Membrane Preparation</th>
<th>Running AA</th>
<th>CG</th>
<th>Level</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHARMM-GUI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>MM, CG</td>
<td><a href="http://www.charmm-gui.org/(19)">http://www.charmm-gui.org/(19)</a></td>
</tr>
<tr>
<td>CABS-flex</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>MM</td>
<td><a href="http://biocomp.chem.uw.edu.pl/CABSflex/(20)">http://biocomp.chem.uw.edu.pl/CABSflex/(20)</a></td>
</tr>
<tr>
<td>locPREFMD</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>MM</td>
<td><a href="http://feig.bch.msu.edu/locpemd/(21)">http://feig.bch.msu.edu/locpemd/(21)</a></td>
</tr>
<tr>
<td>MDWeb</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>MM</td>
<td><a href="http://mmbi.irbbarcelona.org/MDWeb/(22)">http://mmbi.irbbarcelona.org/MDWeb/(22)</a></td>
</tr>
<tr>
<td>PREFMD</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✗</td>
<td>MM</td>
<td><a href="http://feig.bch.msu.edu/prefmd/(23)">http://feig.bch.msu.edu/prefmd/(23)</a></td>
</tr>
<tr>
<td>ProBLM</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>MM</td>
<td><a href="http://compbio.clemson.edu/problm_webserver/(24)">http://compbio.clemson.edu/problm_webserver/(24)</a></td>
</tr>
<tr>
<td>SMOG</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>MM, C-alpha</td>
<td><a href="http://smog-server.org/(25)">http://smog-server.org/(25)</a></td>
</tr>
<tr>
<td>UNRES</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>MM, CG</td>
<td><a href="https://www.unres.pl/web_server/(26)">https://www.unres.pl/web_server/(26)</a></td>
</tr>
<tr>
<td>Vienna-PTM</td>
<td>✓</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>MM</td>
<td><a href="http://vienna-ptm.univie.ac.at/(27)">http://vienna-ptm.univie.ac.at/(27)</a></td>
</tr>
</tbody>
</table>


