



BioScientific Review (BSR)

Volume No.1, Issue No. 1, 2019

ISSN(P): 2663-4198 ISSN(E): 2663-4201

Journal DOI: <https://doi.org/10.32350/BSR>

Issue DOI: <https://doi.org/10.32350/BSR.0101>

Homepage: <https://ssc.umt.edu.pk/Biosci/Home.aspx>

Journal QR Code:



Article: **The Lack of a Centralized and Universal Tool for Input File Generation for Molecular Dynamics Simulations and the Development of such a Tool: A Review**

Muhammad Taimoor Khan

Author(s): Ayesha Ameen

Received: November 8th, 2018

Article Info: Revised: December 23rd, 2018

Accepted for publishing: January 8th, 2019

Article DOI: <https://doi.org/10.32350/BSR.0101.06>

Article QR Code:



Muhammad T Khan

To cite this article: Khan MT, Ameen A. The lack of a centralized and universal tool for input file generation for molecular dynamics simulations and the development of such a tool: A review. *BioSci Rev.* 2019;1(1):48–52.
[Crossref](#)



A publication of the
Department of Life Sciences
School of Science
University of Management and Technology
Lahore

Commentary

The Lack of a Centralized and Universal Tool for Input File Generation for Molecular Dynamics Simulations and the Development of Such a Tool: A Review

Muhammad Taimoor Khan^{1*}, Ayesha Ameen²

¹Department of Computer Sciences, COMSATS University, Islamabad, Pakistan

²Department of Life Sciences, School of Science, University of Management and Technology, Lahore, Pakistan

*Corresponding Author

Muhammad Taimoor Khan

¹Department of Computer Sciences, COMSATS University, Islamabad, Pakistan. Email address: taimoork86@gmail.com

Abstract

This study aims to design and develop a tool that can efficiently generate Molecular Dynamics (MD) simulation input files and systems for various simulation software such as NAMD, Amber and GROMACS collectively using a single platform. Molecular Dynamics simulation is performed in order to understand the structure, dynamics and thermodynamic behavior of biomolecular system. There are a lot of MD simulation software available which are very efficient and fast. However, there are some hurdles in making an input file for simulation purpose, such as the use of command line instructions and the need of using multiple sources in order to generate parameter and topology files. This is a very hectic job for any bio-informatician working in this area and it requires a lot of man hours to develop these files. The main challenge which the scientists encounter working in this field is that there are a lot of different simulation software available online and each one of them has some common features and some uncommon and distinctive features. These facts make input file and system generation a slow and hectic process.

To solve this issue and to minimize the time consumed in the preparation of input files, there is a need to develop a tool which is user friendly and provides a graphical user interface to make an input file.

For this purpose, programming languages like Python and Java will be used during the designing and development process. This will also incorporate all the mathematical equations and calculations that are going on at the back end in the making of input files. Embedding PDB into this tool will also be helpful as users will directly be able to import protein structures from PDB. This software will allow the users to create input files and systems for MD simulations in the shortest possible time.

Keywords: Molecular Dynamics (MD), graphical user interface, PDB: Protein data bank, Charmm, NAMD, VMD, GROMACS, Desmond, Maestro, AMBER, GROMITA, xleap, tleap are names of the frequently used Molecular dynamics simulations software.

1. Introduction

Molecular Dynamic (MD) simulations are important in understanding the structure and functions of proteins, enzymes and other biological macromolecules. It has been possible only due to MD simulations that today we are able to understand the structure, function and interactions among the biological macromolecules which are either difficult or not feasible to understand by any experimental methods. Macromolecules such as proteins were treated as static molecules earlier but MD simulations have enabled us to understand and predict the dynamic nature of proteins. The term MD simulation was first coined in late 1950's when Alder and Wainwright first studied the interactions between hard spheres. The next major advancement occurred in 1964 when Rahman carried out the first simulation with a realistic potential for liquid argon (McCammon, 2002).

With the advent of MD simulations, the level of understanding of protein structure and function has also increased as they play a vital role in explaining the physical basis of the structure and function of the biological macromolecules. MD simulation is a principal tool in the study of biological molecules. It provides us with detailed information about the conformational changes of proteins and nucleic acids and the fluctuations in protein structures. With the passage of time and increasing advancement in science, the algorithmic design and power of computing has also increased enormously. (McCammon, 2011)

In the current era, MD simulation has gained much importance and it has emerged as one of the most powerful computational tools since it has the ability of simulating a huge variety of molecular systems both in and out of thermodynamic equilibrium. A lot of work has been done in this domain during the last few decades. Charmm, Amber, NAMD, GROMACS, Desmond etc. are all the fruits of computer programming that have made MD simulations faster and easier but still there is a lot to be looked into, for example most of the MD simulation software use command line instructions. Hence, users also have to make parameter and topology files from multiple sources for a single simulation. VMD comes with a graphical user interface to generate the input file. Unfortunately, it is time consuming and requires separate parameters and topology files for proteins and ligands. These things make input file and system generation very hectic and lengthy, and sometimes one has to work for days in order to generate a system for the simulation (Allen, 2004).

Commonly used MD simulation software like NAMD, Charm, Amber and GROMACS have the same method of input file and system generation for the simulation to get started, for example in the case of GROMACS the first step to start an MD simulation is to create GROMACS compatible coordinate (.gro) and topology (.top) files and the command line code will look like as follows,

```
pdb2gmx -f 1UBQ.pdb -o protein.gro -p topol.top -ignh-v.
```

However, now there is a graphical user interface present for GROMACS known as Gromita. Gromita's main window is a menu driven interface as well as a button step-by-step layout. Still the problem remains as it was. There is no comprehensive tool which is compatible with most of the MD simulation software for input file and system generation purpose (Sellis, 2009).

Similarly, Amber is also an MD simulation software widely used for research purposes, yet it has also the same limitation. It doesn't come with a GUI to make an input file and system for simulation. This problem is somehow addressed by developing xLeap and tLeap. These programs perform the same function with the difference that xLeap opens in an x-window interface and

presents a very small part in GUI format, whereas tLeap operates from a terminal prompt (no GUI at all). The principal function of these programs is to prepare the AMBER coordinate (inpcrd) and topology (prmtop) files.

CHARMM is also an excessively used MD simulation software for research purposes and it has been provided also with a GUI input generator known as CHARMM-GUI. It is an automated input generation tool for CHARMM to avoid various trivial mistakes in the input. But even after a lot of advancements in this field and despite of the presence of a lot of GUIs for input generation, there is no such comprehensive tool which can at least cover two or three mainly used simulation software and provide an automated input file generation mechanism for them in the form of GUI.

To address these issues a tool will be developed containing an interface sufficient to automate the generation of input files and systems for various MD simulation software. This will reduce the time consumption in generating an input file by avoiding hectic command line instructions and by providing an easy-to-use graphical interface to users. The main purpose of this project is to build a GUI which provides an optimal number of options to users to generate input files as much as possible in a GUI fashion. The automation of this process will help users to avoid various trivial mistakes during input, thus control the quality of calculations with generally accepted standards and well qualified methods and will also allow them to generate input files for various simulation software using a single platform. Such input file and system generation process will help users to modify systems for their research purposes by means of clicking the set of parameters displayed on the GUI of the particular software. Thus the mode of operation of this software will be to generate input and molecular systems based on adjustable parameters with appropriate guides. During each step of input file and system generation short calculations will be performed at the back end of GUI to examine if the generated input and system are properly functional.

2. Statement of Problem

The preparation of input files and systems for running MD simulation is a very time consuming and difficult task. Usually users have to go through several command line instructions and need to use various tools one-by-one to generate an input file and system. Some tools, however, provide graphical user interface to generate input files and systems, such as MOE, Discovery studio visualizer, Maestro and VMD. They are either commercially licensed or need extra files for input file and system generation. The current project, therefore, aims at developing a tool that efficiently generates input files and systems for MD simulation in a GUI fashion without demanding too many files from users, and thus, it will save the time of users.

3. Suggestions of Parameters and Options to be Included in GUI

This software will include parameters and steps which are necessary for the preparation of an input file to be used by an MD simulation software which are as follows,

3.1. Embedding PDB to This Tool

Embedding PDB to this tool will allow the users to directly import the protein structure files from it. This tool will automatically strip water and ligands. The missing residues will also be modelled automatically.

3.2. Addition of Hydrogen Atoms

There will be the option to add hydrogen atoms to our system. The tool will give the option to add either the polar type of hydrogen atoms, that is, OH, NH₂ or all types of Hydrogen atoms, that is, NH₂,CH₂ etc. The tool will enlist both types of hydrogen atoms for the ease of users.

3.3. Ionic concentration

This tool will also give our users an option to add the ions or our users will be able to control the ionic concentration of their molecule as per their requirements. Users will also be able to neutralize the charge on the molecule and may add Na⁺, Cl⁻, K⁺, F⁻ ions as required. The users will have to click the button and the tool will give the option to add any type of ions.

3.4. Water Box addition

This tool will enlist all types of water boxes to be added in an input file, that is, TIP3P, TIP4P etc. So the user just has to select the water force field as desired.

Once all these parameter values are filled in, our system will be prepared and the tool will write the following files.

- Corrected PDB files with water box
- Topology files
- The .rst/.coord files
- Parameter files
- Executable input file with command instructions

This tool will give users the liberty to set the values of following parameters in the executable input file.

- Periodic Boundary Conditions
- Ewald summation method
- Box Grid
- Water Box dimension
- Option of Fixing atom or not
- Ligand preparation
- Forcefield for Ligand
- Timestep value
- Temperature control
- Parameter topology file generation

With these files, the users just need to execute the main input files and the simulation will automatically run on the system.

4. Future aspects

The development of such a tool will prove to be a major step in this direction as it will ease up the simulation process and will pave the road for a new and progressive approach in research. Also, this tool will centralize the simulation software and will work as an anchor so that all simulation software in the future are prepared keeping in view this file generation tool as reference and are made compatible with this tool.

5. Conclusion

The era of information technology and biological sciences is also advancing and developing, hence the conventional methods of research are getting obsolete. With the advent of information technology, more and more tools have been made in order to facilitate the process of research. MD simulation is one of the major fields in biological sciences and a lot of research is going on in this area but the file generation process is a rather difficult and hectic procedure. So, to address this issue it's very much important to develop a tool which should cover all the MD simulation software and bring them all on a single platform.

Competing Interest

None

Funding

None

References

1. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. *Med Res Rev.* 2006;26(5):531–568.
2. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Mol Biol.* 2002;9(9):646–652.
3. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. *BMC Biol.* 2011;9(1):71–79.
4. Anderson JA, Lorenz CD, Travesset A. (2008). General purpose molecular dynamics simulations fully implemented on graphics processing units. *J Comput Phys.* 2008;227(10):5342–5359.
5. Allen MP. Introduction to molecular dynamics simulation. In: Attig N, Binder K, Grubmuller H, Kremer K, eds. *Computational soft matter: From synthetic polymers to proteins* (Lecture Notes). Julich: John von Newmann Institute for Computing; 2004. p. 1–25. NIC series; vol 23.
6. van Gunsteren WF, Dolenc J, Mark AE. Molecular simulation as an aid to experimentalists. *Curr Opin Struct Biol.* 2008;18(2), 149–153.
7. Hünenberger PH, Mark AE, Van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol.* 1995;252(4):492–503.
8. Sittel F, Jain A, Stock G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. *J Chem Phys.* 2014;141(1):014111.
9. Sellis D, Vlachakis D, Vlassi M. Gromita: A fully integrated graphical user interface to gromacs 4. *Bioinf Biol Insights.* 2009;3:99–102.
10. Zhang Q, Zhang W, Li Y, Wang J, Zhang J, Hou, T. (2014). MORT: a powerful foundational library for computational biology and CADD. *Journal of Cheminformatics,* 2014;6(1):36.