Tomasz Makarewicz

Promotor: dr hab. Rajmund Kaźmierkiewicz, prof. UG
Pracownia Symulacji Układów Biomolekularnych

Tworzenie i zastosowanie oprogramowania z graficznym interfejsem użytkownika do przeprowadzania, interpretacji i wizualizacji symulacji dynamiki molekularnej

Creation and application of software, with graphical user interface, that allows users to carry out, analyze and visualize molecular dynamics simulations.

Abstract

Structures of biomolecules are determined by X-ray crystallography, nuclear magnetic resonance or electron cryo-microscopy and are stored digitally in archives such as Protein Data Bank[1]. One of the standard formats of files containing information about the three-dimensional structures of molecules is PDB. Structures saved in this format can be viewed, using most of the software providing molecular visualization, without any problem. One of such programs is PyMOL[2]. It is a popular application written in Python programming language, which allows easy visualization and analysis of structures of biomolecules. Despite the ease of use, PyMOL gives researcher plenty of advanced features for analysis of currently displayed structure. Additionally, it presents an application programmer interface, which allows extension of its features, by plugins prepared specifically for this application. However, an analysis of only the static structure saved in PDB file format is usually not enough to obtain the full knowledge about biomolecule. In a "real" biomolecular system, molecules present dynamic behavior, which has considerable impact on their biological function. One of the methods to obtain knowledge about evolution of molecule conformation in time, is to perform molecular dynamics (MD) simulation starting from the static structure. There are a few software sets, which enable such a calculations. The most popular are: AMBER[3], CHARMM[4] and GROMACS[5]. The last one can be especially considered as a good choice for the researcher, because it is open source and researcher gets unlimited access to its code and can also modify it freely without bearing any costs. However, in contrast to PyMOL, the software for carrying out molecular dynamics simulations (mostly) don't have any graphical user interface (GUI) and the complexity of calculations has its impact on difficulty of using that software. The lack of GUI is justified by the fact, that the long lasting calculations tend to be carried out on the specialized computers, for which it is not allowed to use GUI anyway. Unfortunately, the side effect of such an approach is that it is hard for researcher to start using software for molecular dynamics simulation and researcher needs to achieve relatively high level of knowledge and skills to start using it. On the other hand, contemporary computers have enough power to perform simple molecular dynamics simulation, without the need of specialized machines for this task.

The goal of this dissertation is to propose a new software working as a PyMOL plugin, which allows carrying out the molecular dynamics simulations in the user-friendly environment, using the graphical user interface and visualization of the results in the popular and easy to use PyMOL software. This software, made by myself, is named Dynamics PyMOL Plugin[6].

The Dynamics PyMOL Plugin is written in Python 2 programing language and thanks to that, it is fully compatible with the main PyMOL version. It uses the same Tk library as PyMOL to display graphics, and there is no need for any additional dependencies, that way the proper PyMOL installation is enough to use my plugin. The basic principle underlying the Dynamics PyMOL Plugin action, is to allow researcher to select biomolecule directly from the PyMOL software, and then use the graphical window, designed by myself, to determine options for performing the molecular dynamics simulation and to start the MD procedure, which, after completion, returns the results back to PyMOL. To perform the molecular dynamics simulation run, my application uses the GROMACS package of programs. This requires GROMACS to be present in the system, what shouldn't be a major challenge, because this software is available in repositories of most popular Linux distributions. My Dynamics PyMOL Plugin software is compatible with GROMACS versions 4.x, 5.x and 2016.x. The schematic representation of the software workflow is shown on diagram 1.

In accordance with diagram 1, the communication flow of Dynamics PyMOL Plugin is as follows:

1. After starting PyMOL and loading file containing molecule structure, plugin can be started from the menu bar.

2. In a new window, which appears afterwards (Figure 1), the researcher can set all necessary options required for molecular dynamics simulation. To confirm the choice, the OK button should be clicked.

3. Plugin will use pdb2gmx or x2top GROMACS tools to convert the PDB file loaded in PyMOL, into the trajectory and the topology formats, which can be used for further calculations.

4. Next editconf and gmx solvate will be used to put molecule of interest in solvent and genion will allow addition of determined ions to simulation environment.

5. genrestr tool will restrain fragments of the studied molecule.

6. The main stage of the molecular dynamics simulation will be performed by grompp and mdrun tools.

7. The final result is converted back to PDB format by the trjconv tool and displayed in PyMOL.

8. If the ProDy software is present in the system, it will allow further analysis of the results.

The software is not only easy to use due to application of the graphical user interface and default configuration, which in most cases doesn't require major changes to perform molecular dynamics simulation, but also is a powerful tool for setting GROMACS options. It allows researcher to perform calculation for very specific conditions of simulation. From the main window of the plugin, one can choose most important options, for example which biomolecule should be used for calculations, what part of it, and which empirical force field should be used. All available options are read directly from PyMOL and GROMACS software and the user interface automatically adjusts according to capabilities of versions of installed programs. Additionally, the plugin allows adjusting solvent for specific simulation. First, user has to decide whether to use explicit or implicit solvent model. The implicit solvent is less accurate, but it also requires less computing power and is easier to setup (for example it ignores the boundary conditions). The current GROMACS versions offer the choice of Still[7], HCT (Hawkins-Cramer-Truhlar)[8] and OBC (Onufriev-Bashford-Case)[9] methods for calculating the Born radii. All of them are implemented in my plugin. In case of an explicit solvent, researcher can choose a specific water model. The available water models are determined by selected empirical force field and plugin will automatically show only these water models which are available for the selected force filed. For example for AMBER03 protein, nucleic AMBER94 (Duan et al., J. Comp. Chem. 24, 1999-2012, 2003) force filed there are: TIP3P, TIP4P, TIP4P-Ew, TIP5P, SPC i SPC/E and for the GROMOS96 54a7 force field (Eur. Biophys. J. (2011), 40,, 843-856, DOI: 10.1007/s00249-011-0700-9) there are: SPC i SPC/E. The example showing water model selection is shown on Figure 2.

My application allows also the adjustment of water box in which simulation will take place. Researcher can set geometry, size and density of solvent box. The newest addition is the choice of hydrogen isotope type. In this case the decision can be made between "normal" hydrogen, deuterium or even heavier isotope. My plugin allows setting the electrical charge of the system to 0 by addition of ions. Their concentration can be selected separately. In order to provide more elaborated choice of parameters of molecular dynamics simulation, GROMACS utilizes the text configuration files prepared in .mdp format. My plugin allows edition of default parameters in those files using graphical interface for: energy minimization (em.mdp), the separate molecular dynamics simulation of solvent (pr.mdp) and main molecular dynamics simulation (md.mdp). The example window with edited parameters of md.mdp is shown in Figure 3.

The files can be also prepared separately and placed in the project directory, in this case they will be used instead of the default ones. My software allows also imposing restrains on movement of parts of molecule. This is especially useful, when researcher is only interested in simulation of only a part of the biomolecule, for example its active center. Particularly useful feature is selection of parts of the molecule in PyMOL graphical window and then plugin will inform GROMACS to restrain only those selected atoms. In my opinion, another big feature of the software, worth mentioning, is its integration with ProDy library. If ProDy library is present in the system, it is possible to use it to transform animation, which is the result of molecular dynamics simulation into static snapshot of the molecule with displayed movement vectors. The advantage of such an approach is that result of molecular dynamics simulation can be printed as a single picture and used for example in scientific manuscripts. It is also worth mentioning, that Dynamics PyMOL Plugin is, at the moment, the only way to visualize results of ProDy calculations in PyMOL. The example result is shown in Figure 4.

Additionally, the ProDy library allows visualization of contact maps and cross correlation maps directly in the PyMOL window or as a chart. The example of cross correlation map is shown in chart 1.

Molecular dynamics simulation doesn't need to be performed on a single computer. My Dynamics PyMOL Plugin allows to save and load calculation results and, consequently, it is easy to continue work on the other machine. What is more, in case of simulation failure, my plugin allows easy GROMACS log access, it allows detection of issue and subsequent correction of simulation setup and then repeating simulation or continuing calculation from a break point.

My program – the Dynamics PyMOL Plugin – is developed as a free software and shared under the terms of GPL-3 license. The entire project uses GitHub as a hosting platform and can be accessed through the link: https://github.com/tomaszmakarewicz/Dynamics . This model of software development assures, that, as long as there are people interested in the project, it is still being developed and everyone can use it. Additionally, PyMOL and GROMACS are also open sourced and as a result scientists have access to source code of all of those tools. In the future I am still planing to work on this project and implement among others: distance restrains (current restrains are based on atoms positions), simulations with Martini-like potentials[10], molecular docking simulation, plugin working on a client → server architecture (in such a case, user settings are sent to the server, which performs calculations and sends the results back to "client" computer). However, the real direction of future development of the software will be dictated by its users and their needs. Currently project accepts bug reports, wishes and even code commits in a formalized issue tracker system. As a result, for example, code donation from 11.04.2017 by Ajit Bikram Datta from Department of Biochemistry, Bose Institute, India implements GROMACS module genion and the option to set hydrogen mass, code donation from 31.05.2017 by Manish Sud from MayaChemTools project implements Microsoft Windows support. As a result, I am (mostly) planing to focus on project organization, solving bugs reported by the users and contacting developers, who wants to contribute another features to the project. One of the most important step in this direction will be incorporation of continuous integration system.

After six years of development, my software – the Dynamics PyMOL Plugin grew up from "nothing" to a big project used worldwide, whose development is reported in scientific journals. Among known users of this software are people from such institutions as: Bowdoin College, KU Leuven, Harvard Medical School, NCBI. My project received code donations from institutions: Bose Institute, MayaChemTools. At the moment of writing, my project has almost three thousands lines of code added by more than two hundred forty accepted commits. This software supports Linux, Windows and macOS operating systems together with multiple GROMACS versions. My plugin is available in the main Gentoo Linux distribution repositories as well as in Ubuntu ppa repositories.

I consider Dynamics PyMOL Plugin as a software, which fulfills its destiny of allowing broader access to molecular dynamics simulations for the researchers by the easy to use graphical user interface, default setup of calculation environment and easy access to GROMACS logs. Additionally, my software

allows broad configuration changes of the molecular dynamics simulation conditions and therefore it is suitable even for researchers knowing well a command line tools and requiring access to lots of options. This software is particularly suited for simulation of small biomolecular systems, that doesn't require large computing powers and it can be used for initial checking if simulation conditions are correct, before sending calculations to the remote server. I hope, that the Dynamics PyMOL Plugin project will still be developed, its usability will rise and as a result its usage, among researchers and developers, will grow as well. I hope, that my software will flatten the learning curve for the researchers planning the use of molecular dynamics simulations and as a consequence the interest in this field of science will grow.

The effects of my work are: the described software and two scientific publications.