The Journal of Open Source Software

pyrexMD: Workflow-Orientated Python Package for Replica Exchange Molecular Dynamics

Arthur Voronin[1, 2] **and Alexander Schug***[3, 4]

1 Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany **2** Department of Physics, Karlsruhe Institute of Technology, Karlsruhe, Germany **3** Institute for Advanced Simulation, Jülich Supercomputing Center, Jülich, Germany **4** Faculty of Biology, University of Duisburg-Essen, Duisburg, Germany

DOI: 10.21105/joss.03325

Software

- Review
- Repository
- Archive

Editor: Jeff Gostick

Reviewers:

- @janash
- @rosecers

Submitted: 14 May 2021
Published: 07 July 2021

Summary

Proteins are complex biomolecules which fulfill a wide range of critical tasks in living organisms. Studying and understanding their structure, function, and dynamics is essential for life sciences and can be applied for, e.g., disease control or advanced drug design. Molecular dynamics (MD) is a computational method relying on physical models to simulate biomolecular systems. Movements of all atoms can be 'viewed' like a movie and analyzed to improve the understanding of specific interactions or complement experimental measurements. Replica Exchange (REX) is a powerful method used to enhance the sampling of protein conformations and generates large amounts of data.

`pyrexMD` is designed as an interactive 'all-purpose' toolkit for research projects which rely on (contact-guided) Replica Exchange Molecular Dynamics using `GROMACS`. Due to its workflow-orientated design, it is possible to rapidly create whole setup or analysis workflows, thereby significantly enhancing productivity and reducing the time spent at various stages of the project.

Theoretical background

Timescales of various biomolecular interactions of interest, such as protein folding, conformation transitions, or ligand binding, are typically in the order of µs to s. MD simulations, however, operate in 1-2 fs steps, which makes in-silico studies of proteins computationally demanding. Besides, observations of native configurations during MD simulations are often not guaranteed and proteins can become trapped in certain conformations. One possibility to overcome this problem is to utilize an enhanced sampling technique such as REX (Sugita & Okamoto, 1999; Zhang et al., 2005). REX simulates N non-interacting copies ("replicas") of a system at different temperatures T_i. At certain time intervals adjacent replicas can be exchanged which leads to a walk in temperature space. REX is therefore suited to obtain physically meaningful structure ensembles at specific temperatures. Based on the chosen temperature range and distribution is is also possible to obtain native-like conformations within a single run. Depending on the research goal, it is beneficial to integrate additional theoretically (Morcos et al., 2011) or experimentally derived (Perilla et al., 2017) biases into REX simulations to restrict the sampling space and thus effectively lower computational costs.

*al.schug@fz-juelich.de

The Journal of Open Source Software

Statement of need

In particular analyzing research studies relying on REX can become quite arduous and time consuming. REX simulations usually not only require knowledge of various program tools but also consist of many individual steps ranging from simulation setup and pre-processing over testing and simulation-monitoring to post-processing and data analyses. Furthermore, REX can generate terabytes of data and requires in particular a systematic handling of I/O.

One of the most used software packages for MD is `Gromacs` (Van Der Spoel et al., 2005), a free open source solution enabling the user to choose from many different force fields such as GROMOS (Schmid et al., 2011), AMBER (Wang et al., 2004), CHARMM (MacKerell Jr et al., 2000), or OPLS (Jorgensen et al., 1996). The core functionality of `Gromacs` can be extended by plug-ins such as `PLUMED` (Bonomi et al., 2009; Tribello et al., 2014) or `SSAGES` (Sidky et al., 2018). They implement additional algorithms and enhanced sampling methods which interact during the MD simulation itself or can give access to user-defined collective variables enabling new types of analyses.

`pyrexMD` on the other hand focuses on improvements during the simulation setup or post-simulation analyses. It provides efficient and robust methods for setting up optimized (contact-guided) MD or REX MD simulations. Furthermore it offers many different structure analysis and comparison functions to explore the large I/O sets generated by REX. `pyrexMD` is designed as an interactive 'all-purpose' toolkit which should be used in jupyter notebooks. The provided functions primarily combine many individual steps into workflow-orientated functions which simplify and partially automate large tasks. This allows the user to rapidly create whole setups or analysis workflows by using just a few commands which greatly increases the productivity and reduces the time spent at various stages of the project. Examples of currently available functions include:

- set up systems for normal MD or REX MD simulations
- integration of bias contacts and bias potentials
- topology comparison and consistency checks across different systems or replicas
- trajectory viewer and interactive plots
- wide range of analyses functions related to structure comparison (e.g. contact maps, Qvalues, global distance test, local accuracy, dihedrals, cluster analyses, etc.)
- functions that calculate meaningful measurements such as RMSD, Qvalues, etc. are coupled to plot or logging functions

`pyrexMD` efficiently integrates and extends the following popular MD-related python packages:

- `MDAnalysis` (Gowers et al., 2016; Michaud-Agrawal et al., 2011),
- `nglview` (Nguyen et al., 2018),
- `GromacsWrapper` (Beckstein et al., 2019).

By covering various important aspects, `pyrexMD` allows to execute the whole project from start to finish without switching to other programs which unnecessarily interrupts the workflow and often requires know-how of different command line syntaxes. Alongside the workflow-orientated functions, it also adds a variety of useful general functions and quality of life improvements, such as an integrated trajectory viewer, interactive figures linked to a trajectory or creation of multi-panel figures from saved .pickle files. With `pyrexMD`, it becomes especially easy to create, share and reproduce research results or transfer the work on other target structures of interest. Furthermore, it lowers the technical boundaries for newcomers who want to do research utilizing REX for enhanced sampling.

Example applications

pyrexMD was initially developed during the work of (Voronin et al., 2020). It is currently applied in ongoing REX studies about protein and RNA structure refinement.

Fig. 1 exemplarily shows the application of the trajectory viewer with an interactive plot. Fig. 2 displays a true positive rate analysis of predicted bias contacts based on the number of ranked contacts. Fig. 3 shows the local accuracy of conformations based on a global distance test for models generated during a REX study.

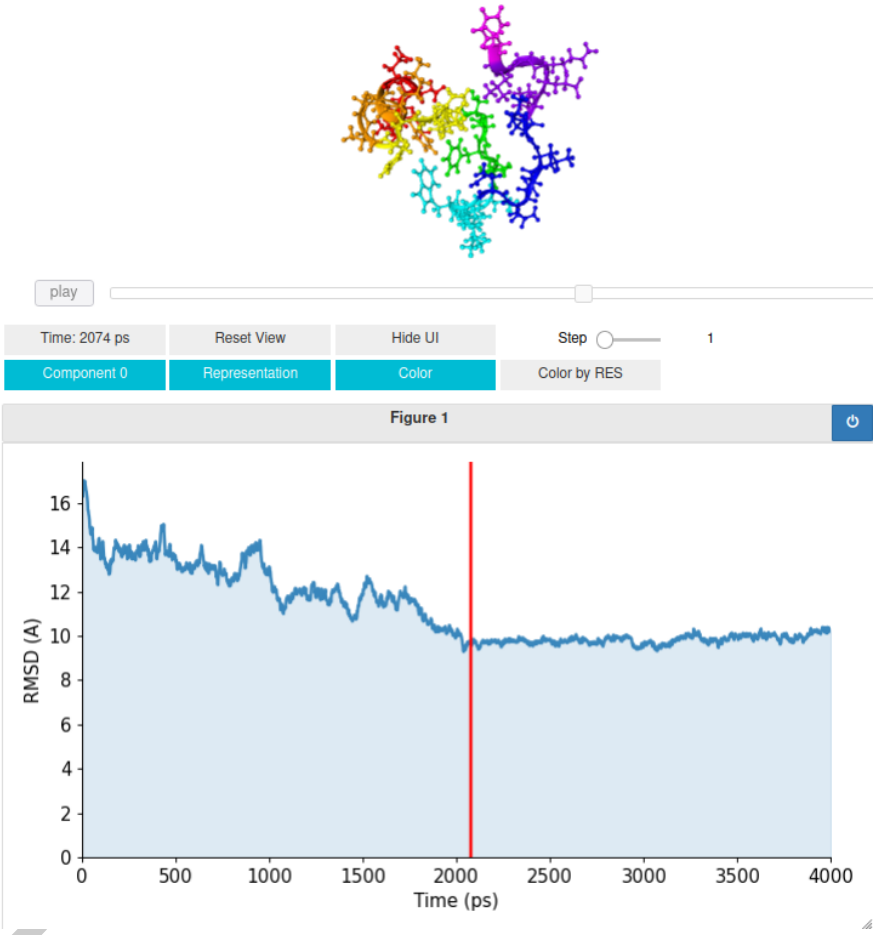

Figure 1: Trajectory viewer on top which is linked to the interactive plot (here RMSD) at the bottom. It is possible to quickly inspect conformations at specific values by interacting with the graph itself (e.g. via ctrl-click) in order to get additional valueable information accessible through the trajectory viewer.

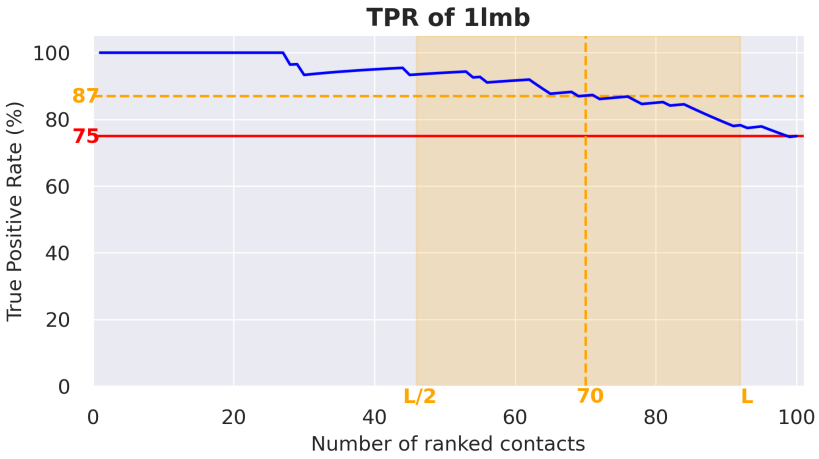

Figure 2: Analysis of the true positive rate (TPR) for bias contacts using `pyrexMD`. The figure shows the TPR of considered bias contacts in addition to other relevant value guidelines according to (Voronin et al., 2020) such as a minimal TPR threshold of 75% in red and a suggested optimal number of contacts in orange between L/2 and L contacts, vizualized by the orange region.

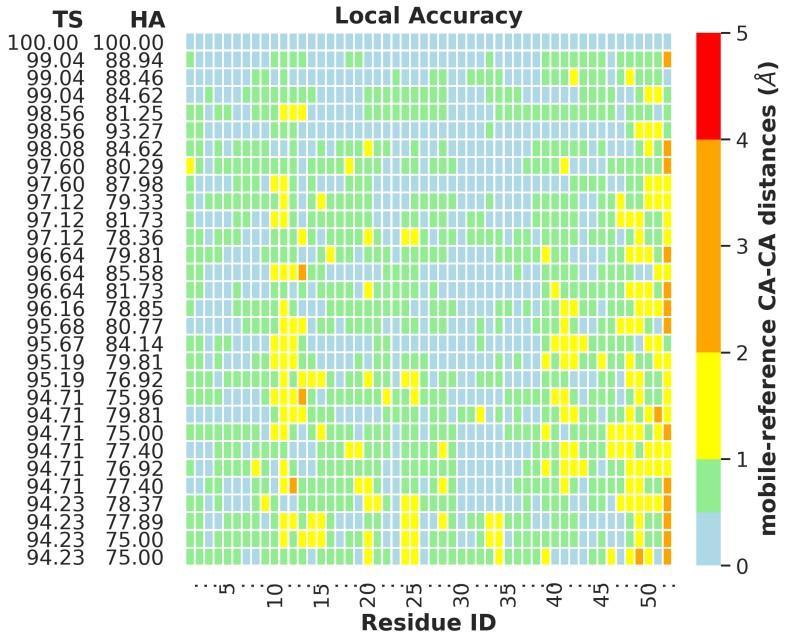

Figure 3: Local accuracy of REX-generated models based on global distance test scores. The figure gives a clear representation of how good each part of the model is refined compared to a reference structure. Each residue is color-coded to represent the CA-CA distance between the model and reference structure after fitting. Corresponding global distance test scores are shown on the left side.

Availability

`pyrexMD` can be downloaded from https://github.com/KIT-MBS/pyrexMD under the MIT license. Both online documentation and quick guide can be accessed via https://kit-mbs.github.io/pyrexMD

Acknowledgments

We want to thank all developers and contributors of MDAnalysis as the backbone of pyrexMD allowing us to parse and analyze MD-related data files.

This work is supported by the Helmholtz Association Initiative and Networking Fund under project number ZT-I-0003.

References

Beckstein, O., Lacerda, P., Domański, J., Dotson, D., Heavey, T., White, A., Gowers, R., Linke, M., Kenney, I., Cody, Fan, S., Somogyi, A., denniej0-2, Loche, P., Mohebifar, M., & Berg, A. (2019). *Becksteinlab/GromacsWrapper: Release 0.8.0* (release-0.8.0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.2654393

Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A., & others. (2009). PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Computer Physics Communications*, *180*(10), 1961–1972. https://doi.org/10.1016/j.cpc.2009.05.011

Gowers, Richard J., Linke, Max, Barnoud, Jonathan, Reddy, Tyler J. E., Melo, Manuel N., Seyler, Sean L., Domański, Jan, Dotson, David L., Buchoux, Sébastien, Kenney, Ian M., & Beckstein, Oliver. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Sebastian Benthall & Scott Rostrup (Eds.), *Proceedings of the 15th Python in Science Conference* (pp. 98–105). https://doi.org/10.25080/Majora-629e541a-00e

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(45), 11225–11236. https://doi.org/10.1021/ja9621760

MacKerell Jr, A. D., Banavali, N., & Foloppe, N. (2000). Development and current status of the CHARMM force field for nucleic acids. *Biopolymers: Original Research on Biomolecules*, *56*(4), 257–265. https://doi.org/10.1002/1097-0282(2000)56:4

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, *32*(10), 2319–2327. https://doi.org/10.1002/jcc.21787

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., & Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, *108*(49), E1293–E1301. https://doi.org/10.1073/pnas.1111471108

Nguyen, H., Case, D. A., & Rose, A. S. (2018). NGLview–interactive molecular graphics for jupyter notebooks. *Bioinformatics*, *34*(7), 1241–1242. https://doi.org/10.1093/bioinformatics/btx789

Perilla, J. R., Zhao, G., Lu, M., Ning, J., Hou, G., Byeon, I.-J. L., Gronenborn, A. M., Polenova, T., & Zhang, P. (2017). CryoEM structure refinement by integrating NMR chemical shifts with molecular dynamics simulations. *The Journal of Physical Chemistry B*, *121*(15), 3853–3863. https://doi.org/10.1021/acs.jpcb.6b13105

Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., & Gunsteren, W. F. van. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal*, *40*(7), 843–856. https://doi.org/10.1007/s00249-011-0700-9

Sidky, H., Colón, Y. J., Helfferich, J., Sikora, B. J., Bezik, C., Chu, W., Giberti, F., Guo, A. Z., Jiang, X., Lequieu, J., & others. (2018). SSAGES: Software suite for advanced general ensemble simulations. *The Journal of Chemical Physics*, *148*(4), 044104. https://doi.org/10.1063/1.5008853

Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, *314*(1-2), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., & Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Computer Physics Communications*, *185*(2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, *26*(16), 1701–1718. https://doi.org/10.1002/jcc.20291

Voronin, A., Weiel, M., & Schug, A. (2020). Including residual contact information into replica-exchange MD simulations significantly enriches native-like conformations. *PloS One*, *15*(11), e0242072. https://doi.org/10.1371/journal.pone.0242072

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174. https://doi.org/10.1002/jcc.20035

Zhang, W., Wu, C., & Duan, Y. (2005). Convergence of replica exchange molecular dynamics. *The Journal of Chemical Physics*, *123*(15), 154105. https://doi.org/10.1063/1.2056540