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
- @jarvist

Submitted: 14 May 2021
Published: 01 December 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 a Python package that is mainly designed for research projects which

- use (contact-guided) REX MD or (contact-guided) MD
- or focus on structure analyses and comparison.

It has three main goals:

1. Interactive 'all-purpose' environment. By including various modified `GROMACS` and `MDAnalysis` Python bindings, this package provides a comprehensive `Jupyter` notebooks based environment to design, run, and analyze MD simulation projects from beginning to end.
2. Data visualization is important. In `pyrexMD`, most analysis functions for calculating useful quantities, such as root-mean-square deviation (RMSD), Q values, contact distances, etc., can generate specialized figures in the same step by passing the keyword argument `plot=True`.
3. User-friendly and simple application. Where possible, the provided functions combine individual steps into comprehensive workflows with additional automation features. It is possible to rapidly create whole setup or structure-analysis workflows within a few commands, thereby significantly enhancing productivity and reducing the time spent at various stages of the project.

With `pyrexMD`, it becomes straightforward to create, share, and reproduce research results or transfer the work to other biomolecular structures of interest. Furthermore, it lowers the technical barrier for non-specialists who want to use REX for enhanced sampling.

*al.schug@fz-juelich.de

Theoretical background

Biomolecular function is often accompanied by slow conformational transitions, typically in the order of µs to s. MD simulations, however, integrate over time in 1-2 fs steps, which makes in-silico studies of proteins computationally demanding. This can lead to incomplete sampling of conformational space as, e.g., proteins can become trapped in specific conformations. One possibility to overcome this problem is to use enhanced-sampling techniques 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. After predefined time intervals, adjacent replicas can be exchanged which leads to a walk in temperature space, speeding up sampling while conserving thermodynamic properties. REX is therefore suited to obtain physically meaningful ensembles of a biomolecular structure at specific temperatures. Based on the chosen temperature range and distribution, native-like conformations can be obtained within a single run. Depending on the research goal, it is beneficial to integrate additional theoretically (Schug et al., 2009) or experimentally derived (Perilla et al., 2017) biases into REX simulations to restrict the sampling space and thus effectively lower computational costs.

Statement of need

Analyzing simulation studies using REX manually is extremely 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 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 providing many different force fields, such as GROMOS (Schmid et al., 2011), AMBER (Wang et al., 2004), CHARMM (Bjelkmar et al., 2010), 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). Such plug-ins implement additional algorithms and enhanced-sampling methods which interact during the MD simulation itself or can give access to user-defined collective variables for new types of analyses.

`pyrexMD` on the other hand focuses on facilitating, assisting, and automating the simulation setup and post-simulation analyses. It provides efficient and robust methods for setting up optimized (contact-guided) REX MD or MD simulations. Furthermore, it offers many intuitive and user-friendly structure analyses and comparison functions to explore the large I/O sets generated by REX.

Examples of currently available functions include:

- setup of systems for MD or REX MD simulations
- integration of bias contacts and bias potentials
- topology comparison functions for consistency checks across different systems or replicas
- trajectory viewer and interactive plots
- wide range of functions related to structure analyses and comparison (e.g. contact maps/distances, RMSD, Q values, global distance test, local accuracy, dihedrals, cluster analyses, etc.)
- easy and interactive data visualization
- automation features and default-parameter switches

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

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

By covering various important aspects, `pyrexMD` allows to execute the whole project from beginning to end without switching to other programs which unnecessarily interrupts the workflow and often requires know-how of different command-line syntaxes. Alongside many workflow-orientated functions, it also adds a variety of useful general functions and workload-reducing improvements, such as an integrated trajectory viewer, interactive figures linked to a trajectory or generation of multi-panel figures from saved .pickle files to reuse individual or old figures without requiring the explicit data set.

Example applications

`pyrexMD` was initially developed in the course of (Voronin et al., 2020). Currently, it is successfully applied in ongoing REX studies on protein and RNA structure refinement.

Figs. 1-3 exemplarily show a small selection of possible data visualizations after performing analyses with `pyrexMD`. Fig. 1 displays the application of the trajectory viewer with an interactive plot. Fig. 2 shows a true-positive-rate analysis of predicted bias contacts which are considered for a contact-guided REX simulation. Fig. 3 visualizes the local accuracy of conformations based on a global distance test for models obtained from a REX study.

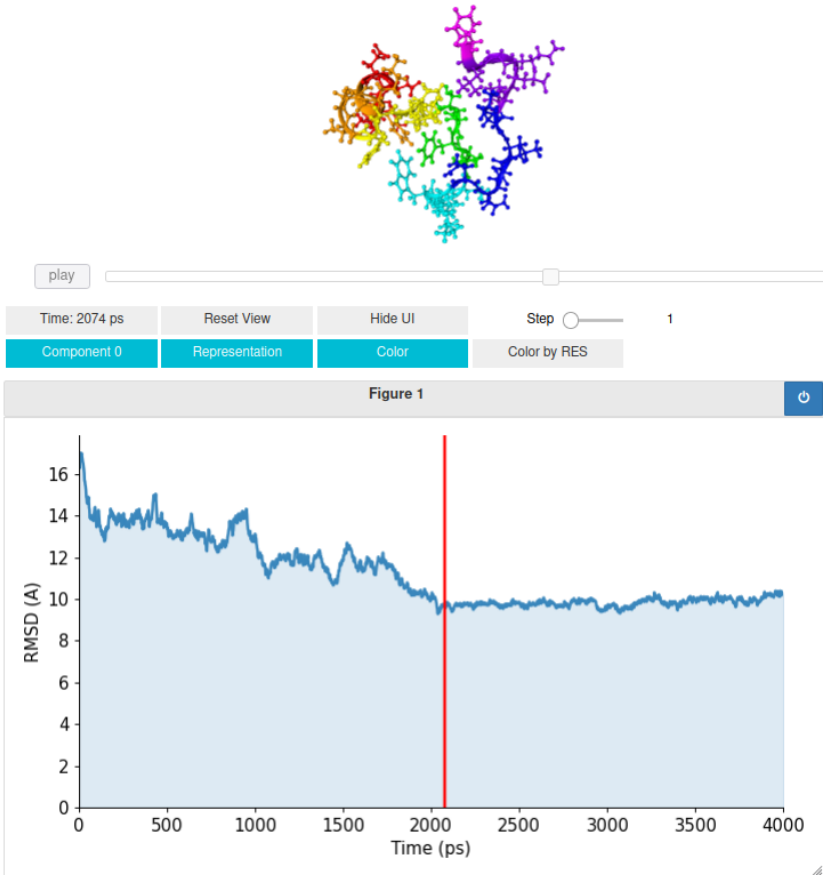

Figure 1: Trajectory viewer (top) which is linked to an interactive plot (here RMSD, bottom). Conformations at specific values can be quickly inspected by interacting with the graph itself (e.g. via ctrl-click), thus making additional valuable information accessible through the trajectory viewer.

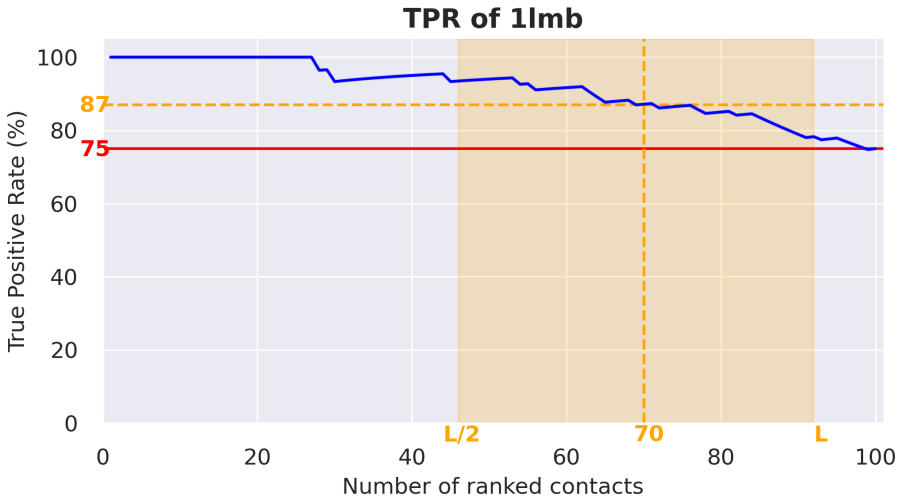

Figure 2: Analysis of the true positive rate (TPR) for bias contacts with `pyrexMD`. The figure exemplarily shows the TPR of the considered bias contacts together with other relevant value guidelines for contact-guided REX (Voronin et al., 2020), such as a minimal TPR threshold of 75% (red) and a suggested optimal number of contacts between L/2 and L (orange), where L denotes the biomolecular sequence length.

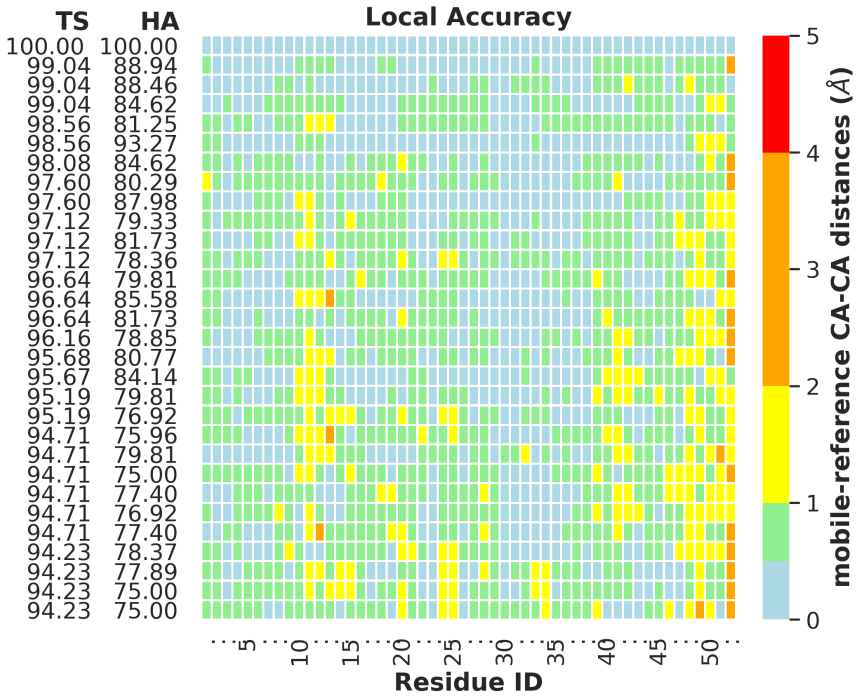

Figure 3: Local accuracy of REX-generated protein models sorted by GDT scores. The so-called global distance test (GDT) is a method for structure evaluation similar to the root-mean-square deviation (RMSD). This figure clearly shows how good each model part is refined compared to a reference structure. Each residue is color-coded to represent the CA-CA distance between the model and reference structure after fitting. The two corresponding GDT score variants Total Score (TS) and High Accuracy (HA) are shown on the left side.

Availability

pyrexMD is free and open source. It is published under the MIT license. You can download the package at https://github.com/KIT-MBS/pyrexMD. 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. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). This work was also performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

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

Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B., & Lindahl, E. (2010). Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation*, *6*(2), 459–466. https://doi.org/10.1021/ct900549r

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

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

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

Schug, A., Weigt, M., Onuchic, J. N., Hwa, T., & Szurmant, H. (2009). High-resolution protein complexes from integrating genomic information with molecular simulation. *Proceedings of the National Academy of Sciences*, *106*(52), 22124–22129. https://doi.org/10.1073/pnas.0912100106

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