

TinkerModellor: An Efficient Tool for Building Biological Systems in Tinker Simulations

Xujian Wang^{1,2,3}, Haodong Liu³, Jiahuang Li³, and Wanlu Li^{1,2}

¹ Department of NanoEngineering, University of California San Diego, CA 92093 ² Program of Materials Science and Engineering, University of California San Diego, CA 92093 ³ School of Biopharmacy, China Pharmaceutical University, China ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [Open Journals](#)

Reviewers:

- [@openjournals](#)

Submitted: 01 January 1970

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

TinkerModellor (TKM) serves as a versatile biological system construction tool designed to create intricate virtual biological systems for molecular dynamics within the Tinker Simulation Program. Its primary function involves the input generation of complex systems compatible with molecular simulation software. TKM possesses the capability to convert various specific formats, such as crd/top in Amber, gro/top in GROMACS, and crd/psf in CHARMM, into the Tinker format (Tinker XYZ). Moreover, TKM offers a user-friendly and concise approach, functioning as an independent script for convenient usability. Additionally, it provides users with several distinct modules to construct personalized workflows, ensuring flexibility and ease of use. Essentially, TKM empowers users to model or simulate within one software and seamlessly transition to Tinker for analysis or further simulation, effectively harnessing separate functionalities concurrently.

Statement of need

Molecular dynamic simulation technology employs advanced computational algorithms to replicate the dynamic behavior of biomolecules or materials at the atomic and molecular scale. This virtual laboratory enables researchers to observe and analyze the intricate interactions and movements of individual atoms, shedding light on phenomena such as protein folding, chemical reactions, and molecular motion in different environments. Particularly, this technology finds extensive application in computer-aided drug design (CADD) (Welborn, 2022), enzyme rational design (Zheng & Vaissier Welborn, 2022), and ab initio antibody design (Chowdhury et al., 2018). Notably, vibrational Stark effect spectroscopy experiments (Fried & Boxer, 2017) have demonstrated the existence of electric fields (EF) that potentially play a crucial role in drug and enzyme design (Bhowmick et al., 2017; Chaturvedi et al., 2023; Li & Head-Gordon, 2021; Vaissier Welborn et al., 2023). However, investigating EF necessitates a high-precision polarization force field to accurately model the dynamic response of electrons to external forces and capture the nuanced effects on molecular structures and interactions. According to the best of our knowledge, AMOEBA/AMEOBAPro (Bim & Alexandrova, 2021; Ren et al., 2011; Shi et al., 2013) stands as a pioneering force field in this domain, effectively integrated into the Tinker simulation program. Furthermore, a post-analysis tool for EF has been proposed (Nash et al., 2020), enabling seamless analysis of trajectory files generated by Tinker. Consequently, utilizing Tinker for EF research emerges as a promising method.

Despite its utility, Tinker has limitations in modeling and simulation processes. It faces challenges in building systems with over 10,000 residues and struggles with accurate topology structure identification. Imperfect parallel computing and the inability to harness GPU acceleration technology act as bottlenecks for Tinker-8. Conversely, GROMACS, Amber, and CHARMM boast robust modules for constructing intricate systems. Additionally, all three

support GPU acceleration, potentially enhancing simulation tasks. Aligning the formats of these software with the Tinker format can assist users in creating customized files seamlessly. Furthermore, it integrates Tinker's specialized EF analysis capabilities into these traditional software, potentially expediting drug and enzyme design processes and related endeavors.

Workflow

TinkerModellor primarily comprises two distinct parts: the GMXMolecule Module, the TinkerModellorSystem Module. Notably, TinkerModellor primarily operates based on the GROMACS format input file. While this might present some limitations, we've mitigated this challenge by leveraging ParmEd, a powerful tool capable of converting between GROMACS, Amber, and CHARMM format files seamlessly. Consequently, users of Amber and CHARMM can also utilize TinkerModellor, as we've successfully integrated ParmEd into its functionalities

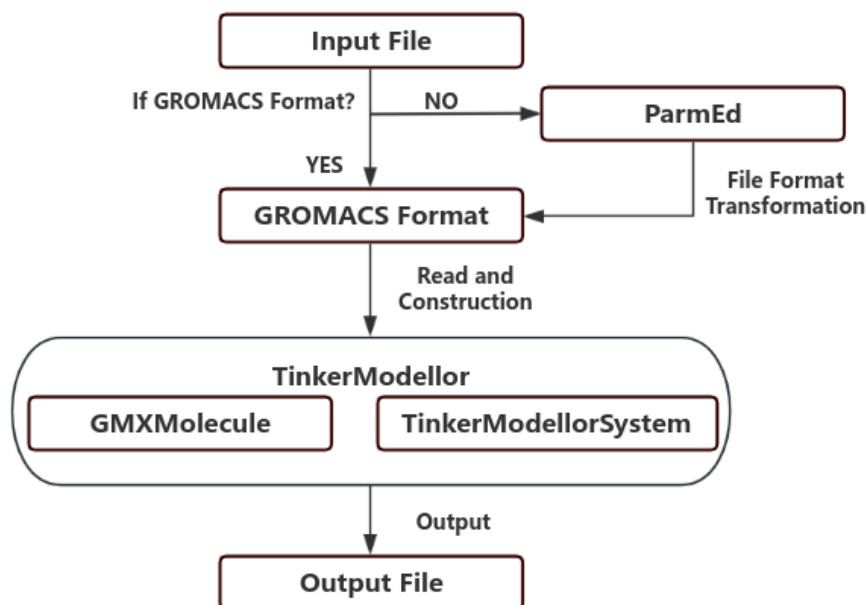


Figure 1. Schematic diagram of TinkerModellor workflow.

The GMXMolecule class is specifically crafted to retain topological details for each molecule type, encompassing bonding relationships and atom types in GROMACS format, among other specifics. Throughout subsequent processes, if a molecule appears multiple times, its topology information will be utilized accordingly, significantly reducing memory usage during system construction. Moreover, the TinkerModellorSystem is devised to encompass the entirety of information requested by Tinker for the system. Essentially, it functions as a data storage class, housing atom types and bonding relationships. However, it also encompasses distinct elements, such as atom indices and coordinates for each atom. These unique details, absent in GMXMolecule, require extraction from coordinate files and subsequent integration with topology information.

The TinkerModellor module necessitates input of both a coordination file and a topology file. It proceeds by invoking functions to read and partition the topology file into distinct molecule types, creating a GMXMolecule object for each type. For instance, a system comprising a protein, a ligand, water, and ions would prompt TinkerModellor to generate GMXMolecule

objects for each corresponding molecule type: protein, ligand, water, and ions. Moreover, the end of the topology file denotes the number of appearance time for each molecule type. TinkerModellor utilizes this information to amalgamate the coordinate file with topology relationships within the GMXMolecule. Subsequently, the topological relationship between coordinates and GMXMolecule is amalgamated and stored in the TinkerModellorSystem object. It's worth noting that within this model, atom types from the topology file within GMXMolecule undergo automatic transformation into the AMOEBA18 force field for simulating polarizable force field systems. Ultimately, TinkerModellor compiles the information from the TinkerModellorSystem object into an output file.

Usage

Command Line Usage

The general usage of the command is as follows:

```
python tkm.py -c coordination_file -p topology_file -out output_file [options]
```

Here is an example of how to use the command:

```
python tkm.py -c my_coordination_file.gro -p my_topology_file.top -o my_output_file.xyz -f G -a True
```

This command will run the TinkerModellor with a GROMACS coordination file my_coordination_file.gro and topology file my_topology_file.top, and it will output the result to my_output_file.xyz. The input file format is set to GROMACS, and the aggressive atomtype matching mode is enabled

Package Usage

```
import tinkermollor as tkm new= tkm() new('gromacs.gro',gromacs.top') new.write_tkm-system('gromacs.xyz')
```

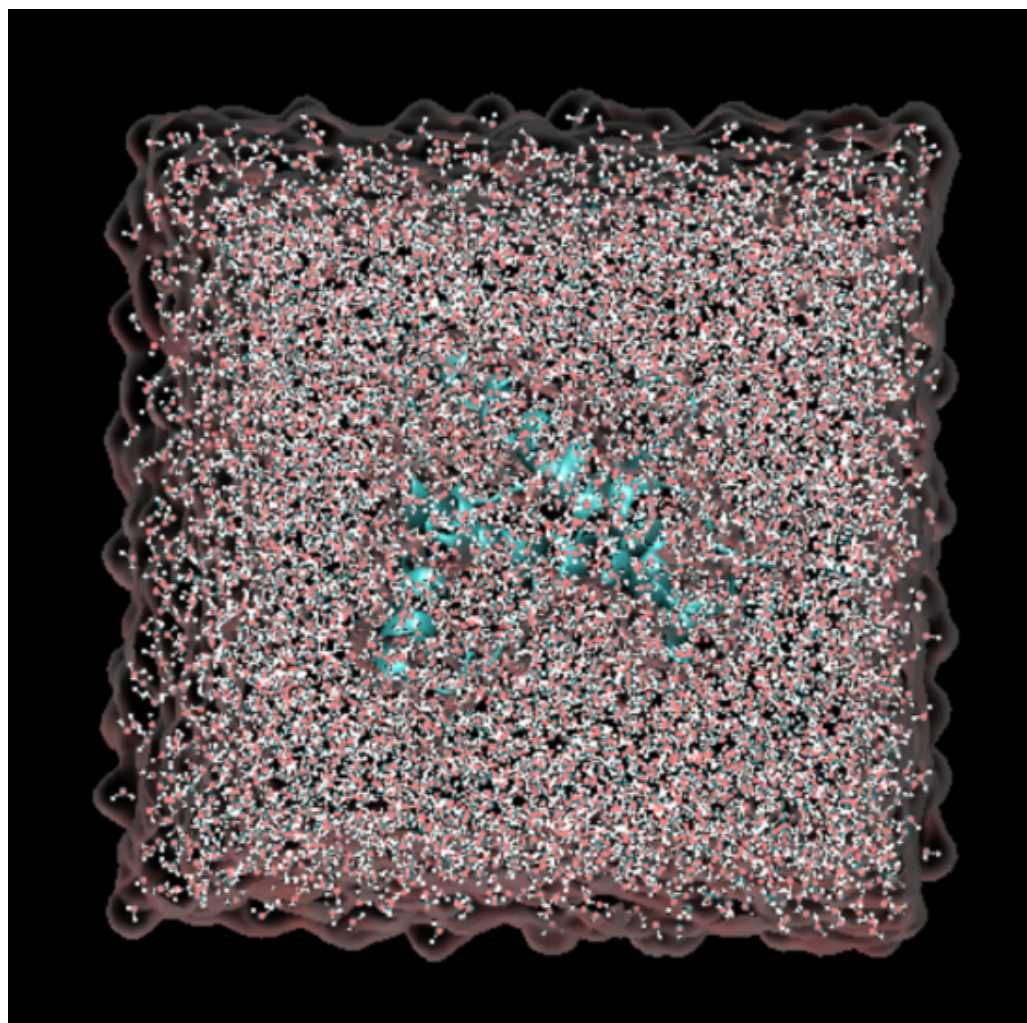


Figure 2. Protein (PDBID:1AG2) in water box. The system is constructed by GROMACS(Abraham et al., 2015) and transformed into Tinker format by TinkerModellor. Mapping by visualization software VMD(Humphrey et al., 1996).

Future Plan

Enhanced Force Field Support: Expand the available force field options to include AMOEBA9(Ren et al., 2011) for protein-ligand simulations, AMOEBANUC17(Zhang et al., 2018) for protein-DNA/RNA simulations, and AMOEBAPRO13(Shi et al., 2013) for more precise biological system simulations. This extension of force field selections will cater to diverse simulation needs across various biological contexts.

Trajectory Transformation: Implementing trajectory adaptation from GROMACS, Amber, and CHARMM to convert their trajectories into Tinker's format will enable simulations across different software platforms. This adaptation allows leveraging the computational efficiency of established biomechanical simulation software like GROMACS, while harnessing Tinker's unique analytical functionalities. This approach promotes the synergy between multiple software, enabling each to capitalize on its specific strengths in simulations and analyses.

Availability

TinkerModellor is freely available under the BSD 3-Clause License on GitHub. For installation instructions, module testing, Python API and tutorial please visit the [GitHub website](#).

Acknowledgements

We acknowledge startup funding from the Jacob school of Engineering, UCSD. This work used the computational resources from the Expanse supercomputer at the San Diego Super Computing Center (SDSC), through allocation CHM230035 and CHE230113. We are also grateful for the financial supports of China National Innovation and Entrepreneurship Training Program for Undergraduate (No.202310316093Y).

Conflict of Interest Statement

The authors declare the absence of any conflicts of interest: No author has any financial, personal, professional, or other relationship that affect our objectivity toward this work.

References

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>
- Bhowmick, A., Sharma, S. C., & Head-Gordon, T. (2017). The importance of the scaffold for de novo enzymes: A case study with kemp eliminase. *J Am Chem Soc*, 139(16), 5793–5800. <https://doi.org/10.1021/jacs.6b12265>
- Bim, D., & Alexandrova, A. N. (2021). Local electric fields as a natural switch of heme-iron protein reactivity. *ACS Catal*, 11(11), 6534–6546. <https://doi.org/10.1021/acscatal.1c00687>
- Chaturvedi, S. S., Bim, D., Christov, C. Z., & Alexandrova, A. N. (2023). From random to rational: Improving enzyme design through electric fields, second coordination sphere interactions, and conformational dynamics. *Chem Sci*, 14(40), 10997–11011. <https://doi.org/10.1039/d3sc02982d>
- Chowdhury, R., Allan, M. F., & Maranas, C. D. (2018). OptMAVEN-2.0: De novo design of variable antibody regions against targeted antigen epitopes. *Antibodies (Basel)*, 7(3). <https://doi.org/10.3390/antib7030023>
- Fried, S. D., & Boxer, S. G. (2017). Electric fields and enzyme catalysis. *Annu Rev Biochem*, 86, 387–415. <https://doi.org/10.1146/annurev-biochem-061516-044432>
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14, 33–38. <https://doi.org/10.1088/2516-1075/ac509f>
- Li, W. L., & Head-Gordon, T. (2021). Catalytic principles from natural enzymes and translational design strategies for synthetic catalysts. *ACS Cent Sci*, 7(1), 72–80. <https://doi.org/10.1021/acscentsci.0c01556>
- Nash, J., Barnes, T., & Welborn, V. (2020). ELECTRIC: Electric fields leveraged from multipole expansion calculations in tinker rapid interface code. *Journal of Open Source Software*, 5(54). <https://doi.org/10.21105/joss.02576>

- 149 Ren, P., Wu, C., & Ponder, J. W. (2011). Polarizable atomic multipole-based molecular
150 mechanics for organic molecules. *J Chem Theory Comput*, 7(10), 3143–3161. <https://doi.org/10.1021/ct200304d>
151
- 152 Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W., & Ren, P. (2013). The polarizable
153 atomic multipole-based AMOEBA force field for proteins. *J Chem Theory Comput*, 9(9),
154 4046–4063. <https://doi.org/10.1021/ct4003702>
- 155 Vaissier Welborn, V., Archer, W. R., & Schulz, M. D. (2023). Characterizing ion-polymer inter-
156 actions in aqueous environment with electric fields. *J Chem Inf Model*, 63(7), 2030–2036.
157 <https://doi.org/10.1021/acs.jcim.2c01048>
- 158 Welborn, V. V. (2022). Beyond structural analysis of molecular enzyme-inhibitor interactions.
159 *Electronic Structure*, 4(1). <https://doi.org/10.1088/2516-1075/ac509f>
- 160 Zhang, C., Lu, C., Jing, Z., Wu, C., Piquemal, J. P., Ponder, J. W., & Ren, P. (2018).
161 AMOEBA polarizable atomic multipole force field for nucleic acids. *J Chem Theory*
162 *Comput*, 14(4), 2084–2108. <https://doi.org/10.1021/acs.jctc.7b01169>
- 163 Zheng, Y., & Vaissier Welborn, V. (2022). Tuning the catalytic activity of synthetic enzyme
164 KE15 with DNA. *J Phys Chem B*, 126(18), 3407–3413. <https://doi.org/10.1021/acs.jpcc.2c00765>
165

DRAFT