




CodeEntropy: A Python package for Multiscale Entropy and Structure Quantification from Molecular Dynamics Simulation

Arghya Chakravorty^{1*}, Jas Kalayan^{2*}, and Donald Chung^{3*}

¹ University of Michigan, Ann Arbor, USA ² University of Manchester, United Kingdom ³ Science and Technology Facilities Council, United Kingdom * These authors contributed equally.

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Published: unpublished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

Entropy is a fundamental property of any system that quantifies the structural flexibility of the system and together with energy governs system stability. It is particularly important in biomolecular systems because of their highly flexible and complex nature. Many properties are a delicate balance of entropy versus energy, necessitating the determination of entropy in order to understand stability. Moreover, entropy provides a way to quantify the structural flexibility of a system over all its degrees of freedom. CodeEntropy is a code based on the Multiscale Cell Correlation (MCC) method which is a novel solution to the problems encountered by other methods by providing a single, scalable and general framework applicable to all molecules in the system.

CodeEntropy is a code that calculates entropy with Multiscale Cell Correlation (MCC) method based on molecular simulation trajectories. The code is based on a combination of CodeEntropy ([Chakravorty, 2021](#)) by Dr Arghya Chakravorty and POSEIDON ([Kalayan, 2021](#)) by Jas Kalayan. The code written by Dr Chakravorty accounts for the vibrational and conformational entropy in a multiscale formulation using an Applications Programming Interface that makes it highly customisable. To make CodeEntropy fully applicable to biomolecular systems, the POSEIDON code will calculate the topographical entropy terms for solvents and mixtures. The topology/trajectory parser, atom selector and distance calculation is performed by the MDAnalysis ([Gowers et al., 2016](#); [Michaud-Agrawal et al., 2011](#)) package.

Statement of needs

There are a range of existing methods to calculate entropy from molecular dynamics simulations but they suffer from a number of limitations: they may only work for particular kinds of system or degrees of freedom, they may require additional calculations, they can be difficult to interpret, or they do not scale well to large and complex systems. Some methods only work for water, such as Inhomogeneous Solvation Theory, for liquids such as 2-Phase Thermodynamics, for only some degrees of freedom such as dihedral binning, or for single molecules such as Quasiharmonic Analysis, Normal Mode Analysis or non-parametric methods such as Minimal Spanning Tree or K-Nearest-Neighbours.

Given the widespread use of free-energy calculations and molecular dynamics simulations, there is a large user-community for software to calculate entropy and quantify full structural flexibility of biomolecular systems. Multiscale Cell Correlation (MCC) provides a novel solution to the problems encountered by other methods by providing a single, scalable and general framework applicable to all molecules in the system. It utilises a judicious synthesis of mean-field cell theory and covariance matrices over a range of length scales: First, Correlations are

41 considered between groups of locally connected atoms as in a mean-field cell, and longer-range
42 correlations are accounted for using a coarser representation of the groups, a framework that
43 is scaled to higher length scales. Second, At each length scale, the potential energy surface is
44 discretised into energy wells for translational and rotational motion. These are represented as an
45 average energy well and an energy-well distribution, denoted as vibrational and topographical,
46 respectively. Finally, The decomposition over molecules, length scales, type of motion and
47 energy-well size and distribution provides an exquisite level of detail in explaining the entropies
48 obtained.

49 MCC has been applied by the group of RH to a wide range of systems, namely liquids, aqueous
50 and octanol solutions, host-guest complexes, chemical reactions and large biomolecules such
51 as proteins, DNA and membrane bilayers in aqueous electrolytes.

52 References

- 53 Chakravorty, A. (2021). CodeEntropy. In *GitHub repository*. GitHub. <https://github.com/arghya90/CodeEntropy>
- 54
- 55 Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., Seyler, S., Domański, J., Dotson, D.,
56 Buchoux, S., Kenney, I., & Beckstein, O. (2016). MDAnalysis: A python package for the
57 rapid analysis of molecular dynamics simulations. *Proceedings of the Python in Science*
58 *Conference*. <https://doi.org/10.25080/majora-629e541a-00e>
- 59 Kalayan, J. (2021). PoseidonBeta. In *GitHub repository*. GitHub. <https://github.com/jkalayan/PoseidonBeta>
- 60
- 61 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis:
62 A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational*
63 *Chemistry*, 32(10), 2319–2327. <https://doi.org/10.1002/jcc.21787>