

Biotite: A bioinformatics framework for sequence and structure data analysis

Author : Muniba Faiza

Categories : [Sequence Analysis](#), [Softwares](#), [Structural Bioinformatics](#)

Date : October 4, 2018

Sequence and structural data in bioinformatics are ever-increasing and the need for its analysis is ever-demanding likewise. As bioinformaticians analyze the data with their keen knowledge and reach important conclusions, similarly, bioinformaticists provide with the enhanced and advanced tools and software for data analysis. There are some computational biology frameworks available for the structural data analysis of molecular dynamics simulation such as MDAnalysis [1] and MDTraj [2]. A new framework has been introduced known as Biotite, a Python package used to represent sequence and structure data [3].

The package is open source and freely available at GitHub (<https://github.com/biotite-dev/biotite>). This package is simple to use especially for the beginners in programming and computationally efficient because of the implementation of Numpy and Cython. Biotite consists of four sub packages: *sequence*, *structure*, *databases*, and *application*. The *sequence* and *structure* modules serve for the analysis of sequence and structural data analysis respectively, *database* downloads files from the other databases such as RCSB PDB, and *application* provides interface for external software [3].

The *sequence* subpackage encodes each character of the sequence into a symbol code which is stored in a NumPy ndarray in the sequence object. The nucleotide and protein sequences can be read and written into FASTA format. Besides, sequences can be easily aligned globally [4] and locally [5] using dynamic programming and can be easily visualized according to the similarity percentage.

The *structure* subpackage uses AtomArrayStack to represent multi-model three-dimensional structures of proteins which has a $(m \times n \times 3)$ coordinate ndarray with n number of atoms and m number of models, and easily parse the files in MMTF format [6]. It is also capable of loading trajectories files of molecular dynamics simulation and can measure angles, dihedrals, and distances between the atoms. Besides, users

can easily perform structure superimposition and calculate RMSD, RMSF, and secondary structure assignment.

Biotite is an efficient framework for bioinformatics analyses such as downloading files, reading and writing structural files, and their modification.

For further reading, [click here](#).

References

1. Michaud?Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAAnalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, 32(10), 2319-2327.
2. McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., ... & Pande, V. S. (2015). MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical journal*, 109(8), 1528-1532.
3. Kunzmann P., Hamacher K. (2018) Biotite: a unifying open source computational biology framework in Python. *BMC Bioinformatics*, 19 (1), 346
4. Waterman, M. S. (1981). Identification of common molecular subsequence. *Mol. Biol*, 147, 195-197.
5. Gotoh, O. (1982). An improved algorithm for matching biological sequences. *Journal of molecular biology*, 162(3), 705-708.
6. Bradley, A. R., Rose, A. S., Pavelka, A., Valasatava, Y., Duarte, J. M., Prli?, A., & Rose, P. W. (2017). MMTF—An efficient file format for the transmission, visualization, and analysis of macromolecular structures. *PLoS computational biology*, 13(6), e1005575.

Sharing is caring. Spread the love!

- [Print](#)
- [Email](#)
- [LinkedIn](#)
- [Twitter](#)
- [Facebook](#)
- [Google](#)
-