

# BioPandas: Working with molecular structures in pandas DataFrames

Sebastian Raschka<sup>1</sup>

<sup>1</sup>Michigan State University, East-Lansing, USA

31 May 2017

**Paper DOI:** <http://dx.doi.org/10.21105/joss.00279>

**Software Repository:** <https://github.com/rasbt/biopandas>

**Software Archive:** <http://dx.doi.org/10.5281/zenodo.804030>

## Summary

BioPandas is a Python library that reads molecular structures from 3D-coordinate files, such as PDB (H. M. Berman 2000) (H. Berman, Henrick, and Nakamura 2003) and MOL2, into pandas DataFrames (McKinney and Others 2010) for convenient data analysis and data mining related tasks.

In addition to parsing protein and small molecule data into a data frame format, BioPandas provides additional utility functions for structure analysis. These functions include common computations such as computing the root-mean-squared-deviation between structures and converting protein structures into primary amino acid sequence formats.

Furthermore, useful small-molecule related functions are provided for reading and parsing millions of small molecule structures (from multi-MOL2 files (Tripos 2007)) fast and efficiently in virtual screening applications. Inbuilt functions for filtering molecules by the presence of functional groups and their pair-wise distances to each other make BioPandas a particularly attractive utility library for virtual screening and protein-ligand docking applications.

## References

- Berman, H. M. 2000. "The Protein Data Bank." *Nucleic Acids Research* 28 (1). Oxford University Press: 235–42. doi:10.1093/nar/28.1.235.
- Berman, Helen, Kim Henrick, and Haruki Nakamura. 2003. "Announcing the worldwide protein data bank." *Nature Structural & Molecular Biology* 10 (12). Nature Publishing Group: 980.
- McKinney, Wes, and Others. 2010. "Data structures for statistical computing in python." In *Proceedings of the 9th Python in Science Conference*, 445:51–56. van der Voort S, Millman J.
- Tripos, L. 2007. "Tripos Mol2 File Format." *St. Louis, MO: Tripos*.