

Project RACCOON: Automated construction of PDB files for polymers and polymer peptide conjugates

Moritz L. Obenauer¹✉, Kai N. Spauszus^{1*}, Pol Besenius¹, and Friederike Schmid²

¹ Department of Chemistry, Johannes Gutenberg-University Mainz, Duesbergweg 10-14, D-55128 Mainz, Germany ² Institute of Physics, Johannes Gutenberg University, Staudingerweg 9, 55128 Mainz, Germany ✉ Corresponding author * These authors contributed equally.

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

Software

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

Editor: ✉

Submitted: 10 January 2024

Published: unpublished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](https://creativecommons.org/licenses/by/4.0/))[®].

Summary

The *Project RACCOON* software was developed to automatically generate PDB files for linear polymers, polypeptides, and polymer peptide conjugates. Previously published software cannot easily represent polypeptide strands with converging termini. *Project RACCOON* also offers the advantage of combining multiple modeling resolutions within one system. Although this software was explicitly developed for C2 symmetric polymer peptide conjugates introduced by Otter et al. (2018) with the OPLS-AA/M force field, its application is not limited to this molecule class. (Robertson et al., 2019; Sakae & Okamoto, 2010) PDB files can be created for all polymers, polypeptides with natural and non-natural amino acids, and other macromolecules combined with any force field.

Statement of need

Common software packages for generating the input structures of molecular dynamics simulations such as *polyply* or *CHARMM-GUI* are unsuitable for representing the class of C2 symmetric polymer peptide conjugates. (Grünwald et al., 2022; Jo et al., 2008) This class of molecules is characterized by the convergence of two C-termini in the polypeptide sequences. This represents a non-natural system and has not yet been implemented in the aforementioned software packages. Modeling these sequences is straightforward and achievable for novice users with *Project RACCOON*, as the flipping of N- and C-termini of amino acids is an integral part of the software. Construction of the molecules is carried out in a building block fashion, allowing the automatic import of natural and non-natural amino acid repeating units and synthetic polymers. In addition, this modular approach offers the advantage of combining building blocks with different resolutions (atomistic, united atom, and coarse-grained) in one molecule. This is necessary to describe polymer peptide conjugates on hybrid scales. (Taylor & Jayaraman, 2020) The software presented here is an improved alternative to other options like *polyply* because the generated files are directly available in the PDB standard format, which can be processed and visualized by any cheminformatic software. Until now, the PDB format has been used almost exclusively for proteins in the bioinformatics context. (Westbrook & Fitzgerald, 2003) However, this work shows that extending the PDB files to general macromolecules simplifies and standardizes data processing. This software represents an important step towards standardization and simplification in the field of cheminformatics and thus contributes to the idea of the UNESCO Open Science Initiative. (UNESCO, 2021)

Functionality and Extensibility

New monomers can be quickly and systematically added to the software. Each monomer is assigned various properties such as the number of atoms, linkage points, and the geometry of the monomer building block. This data is saved as a JSON file and can therefore be easily imported, edited, and exported. (Van Rossum & Drake, 2009) The geometry of the monomer building block can be obtained from vacuum energy-minimized structures or crystallographic data. Bonds link all monomers together, which the user can also define, and are only limited by the capacity of the force field used later to account for these bonds. Each atom is also assigned an individual name, which serves as an identifier for the associated force field. Novice users can utilize the automated import feature on molecular building blocks generated with software like Avogadro 2. (Hanwell et al., 2012) Advanced users can import Project RACCOON into their projects and use all functions for further future work, as the software can quickly adapt to their needs. Functionality and stability were extensively verified with unit testing, resulting in a code coverage of over 85%. Project RACCOON can enhance productivity and collaboration for all users since one can build libraries of monomers and share them with others. It should be noted, however, that the geometry is generated from a self-avoiding random walk and has no physical justification. Parameters for this self-avoiding random walk can be chosen by the user to obtain a pseudo phantom chain structure. Energy minimization steps and unrestricted equilibration steps must be performed with suitable software in any case before using the model in molecular dynamics simulations.

Acknowledgements

Funding from the DFG (*Deutsche Forschungsgemeinschaft*) is acknowledged. P.B. and F.S. are members of the GRK 2516 (Project No. 405552959). This work was supported by the German Academic Exchange Service (DAAD) International Study and Training Partnerships (ISAP) program (Grant No. 57460493). M.L.O. received a scholarship to pursue a research internship in the research group of M. Muthukumar, Polymer Science and Engineering, University of Massachusetts, Amherst.

Conflict of Interest

The authors declare no conflict of interest.

References

- Grünewald, F., Alessandri, R., Kroon, P. C., Monticelli, L., Souza, P. C. T., & Marrink, S. J. (2022). Polyppy; a python suite for facilitating simulations of macromolecules and nanomaterials. *Nature Communications*, 13(1), 68. <https://doi.org/10.1038/s41467-021-27627-4>
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1), 1–17.
- Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11), 1859–1865. <https://doi.org/https://doi.org/10.1002/jcc.20945>
- Otter, R., Henke, N. A., Berac, C., Bauer, T., Barz, M., Seiffert, S., & Besenius, P. (2018). Secondary Structure-Driven Hydrogelation Using Foldable Telechelic Polymer–Peptide Conjugates. *Macromolecular Rapid Communications*, 39(17), 1800459. <https://doi.org/10.1002/marc.201800459>

- 82 Robertson, M. J., Qian, Y., Robinson, M. C., Tirado-Rives, J., & Jorgensen, W. L. (2019).
83 Development and Testing of the OPLS-AA/M Force Field for RNA. *Journal of Chemical*
84 *Theory and Computation*, 15(4), 2734–2742. <https://doi.org/10.1021/acs.jctc.9b00054>
- 85 Sakae, Y., & Okamoto, Y. (2010). Optimisation of OPLS-UA force-field parameters for
86 protein systems using protein data bank. *Molecular Simulation*, 36(14), 1148–1156.
87 <https://doi.org/10.1080/08927022.2010.507247>
- 88 Taylor, P. A., & Jayaraman, A. (2020). Molecular Modeling and Simulations of Peptide–Polymer
89 Conjugates. *Annual Review of Chemical and Biomolecular Engineering*, 11(1), 257–276.
90 <https://doi.org/10.1146/annurev-chembioeng-092319-083243>
- 91 UNESCO. (2021). *UNESCO recommendation on open science*. Electronic Paper.
92 <https://doi.org/10.54677/MNMH8546>
- 93 Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
94 ISBN: 1441412697
- 95 Westbrook, J. D., & Fitzgerald, P. (2003). The PDB format, mmCIF, and other data formats.
96 *Methods Biochem Anal*, 44, 161–179.

DRAFT