The Journal of Open Source Software

buildH: Build hydrogen atoms from united-atom molecular dynamics of lipids and calculate the order parameters

Hubert Santuz[1, 2], **Amélie Bacle**[3], **Pierre Poulain**[4], **and Patrick F.J. Fuchs***[5, 6]

1 CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 Rue Pierre et Marie Curie, F-75005 Paris, France **2** Institut de Biologie Physico-Chimique–Fondation Edmond de Rothschild, PSL Research University, Paris, France **3** Laboratoire Coopératif "Lipotoxicity and Channelopathies - ConicMeds," Université de Poitiers, F-86000 Poitiers, France **4** Université de Paris, CNRS, Institut Jacques Monod, F-75006, Paris, France **5** Sorbonne Université, Ecole Normale Supérieure, PSL Research University, CNRS, Laboratoire des Biomolécules (LBM), F-75005 Paris, France **6** Université de Paris, UFR Sciences du Vivant, F-75013 Paris, France

DOI: 10.21105/joss.03521

Software

- Review
- Repository
- Archive

Editor: Richard Gowers

Reviewers:

- @lilyminium
- @ojeda-e
- @blakeaw

Submitted: 08 July 2021
Published: 22 July 2021

Background

Molecular dynamics (MD) simulations of lipids are widely used to understand the complex structure and dynamics of biological or model membranes (Feller, 2000; Lyubartsev & Rabinovich, 2011; Tieleman et al., 1997). They are very complementary to biophysical experiments and have thus become important to get insights at the microscopic scale. Many of them are performed using all-atom (AA) or united-atom (UA) representations. In AA force fields (such as CHARMM36 (Klauda et al., 2010)), all the atoms are considered whereas in UA force fields (such as Berger (Berger et al., 1997)) the aliphatic hydrogen atoms (Hs) are merged to their parent carbon into a larger particle representing a CH, CH2 or CH3 (e.g. a methyl group is represented by a single CH3 particle). In simulations of phospholipids, the use of UA representations allows to divide by almost 3 the number of atoms to simulate because these molecules contain many aliphatic Hs. It thus reduces the computational cost without loosing important chemical details.

MD simulations of lipids are usually validated against experimental data (Klauda et al., 2010) or used to help interpret experiments (Feller, 2007). One type of experiment which is often used for that is ^2H NMR. In this type of experiment, aliphatic Hs atoms are replaced by deuterons. ^2H NMR allows one to measure the order parameter of a given C-H bond (where the H is replaced by a deuteron):

$$S_{CH} = \frac{1}{2}\left\langle 3cos^2(\theta) - 1 \right\rangle$$

where θ is the angle between the C-H bond and the magnetic field. The symbol $\langle ... \rangle$ means averaging over time and molecules.

This order parameter is useful because it is directly related to the flexibility of the given C-H bond. It describes the amount of possible orientations visited by the C-H bond, varies between 1 and -0.5 and is unitless. Values close to 0 mean high mobility, when it goes away from 0 (towards negative or positive values), it means the C-H bond gets less mobile. Although it varies theoretically between 1 and -0.5, its absolute value $|S_{CH}|$ is often reported because

*corresponding author

its sign is usually difficult to measure experimentally (Ollila, 2014). In MD simulations, θ is the angle between the C-H bond and the normal to the membrane (usually the z axis). For AA simulations, S_{CH} is trivial to calculate. However, it is more difficult for UA simulations since the aliphatic Hs are not present. There are two main strategies to compute S_{CH} from UA simulations (Piggot et al., 2017): i) expressing S_{CH} as a function of the coordinates of other atoms (Douliez et al., 1995), ii) reconstructing Hs coordinates and calculate S_{CH} as in AA simulations. The trend in the last years was more to use strategy ii), such as in the NMRlipids project (Antila et al., 2019; Bacle et al., 2021; Botan et al., 2015; Catte et al., 2016). NMRlipids is an open science project which uses MD simulations and experimental S_{CH} with the goal of improving lipid force fields or making fundamental research on the structure and dynamics of lipid membranes.

Statement of Need

Reconstructing Hs from the heavy atom coordinates can be done using standard geometric rules respecting stereochemistry. In the first NMRlipids project (Botan et al., 2015), H reconstruction was performed using a program called `g_protonate` from the software GROMACS version 3.* (Berendsen et al., 1995). However, `g_protonate` has been removed from GROMACS version 4.* or higher. So, there was a need to find other solutions. Currently, there are many programs in the field of chemoinformatics that are able to build Hs such as OpenBabel (O'Boyle et al., 2011) or other proprietary softwares. It is also possible to use `pdb2gmx` from the GROMACS software (Abraham et al., 2015) to build Hs, but it is not initially intended for that since its main purpose is to build a topology. Many of these solutions remain workarounds where one uses a sophisticated software for just doing one basic task. Using these open or proprietary software usually has a slow learning curve. Moreover, it is not always easy to use them in the Unix command line when they have complicated graphical interfaces, thus complicating their use in automatic pipelines of trajectory analyses. Here, we propose the `buildH` software to meet this need. `buildH` is very light and usable in the Unix command line or as a Python module, making it a tool of choice to integrate in analyses pipelines or Jupyter notebooks. It can be easily extended or customized if needed. `buildH` is currently widely used in the NMRlipids project IVb dealing with the conformational plasticity of lipid headgroups in cellular membranes and protein-lipid complexes (Bacle et al., 2021). In addition, it is planned to use `buildH` in the next NMRlipids project dealing with a databank containing MD trajectories of lipids (Kiirikki & Ollila, 2021).

Overview

`buildH` is a Python software that enables automatic analyses of order parameter calculations from UA trajectories of lipids. The software has the following features:

- It reads a single structure or a trajectory of lipids in a UA representation.
- It reconstructs the aliphatic Hs using standard geometric rules.
- From the reconstructed Hs, it calculates and outputs the order parameters on each requested C-H bond.
- Optionally, it outputs a structure (in pdb format) and a trajectory (in xtc format) with all reconstructed Hs.

Beyond order parameter calculations, the trajectory with Hs can be used for any further analyses (e.g. precise molecular volume calculation).

`buildH` has been natively developed for a use in the Unix command line. It possesses a minimum number of options, making it easy to use. It is also possible to utilize it as a Python module, which may be convenient in some cases.

To reconstruct H atoms, `buildH` uses standard geometric rules. These rules require so-called *helper* atoms. For example, the reconstruction of the two Hs of a CH2 on carbon C_i, requires two helpers which are C_{i-1} and C_{i+1}, that is, the two neighbors of C_i in the chain (note that helpers can also be other heavy atoms such as oxygen or nitrogen). The list of helpers used for the reconstruction of each H is written in a json file. Many json files are already present on the `buildH` repository representing the major lipids: Phoshphatidylcholine (PC), Phoshphatidylethanolamine (PE), Phoshphatidylglycerol (PG) for the polar heads and palmitoyl, myristoyl, oleoyl for the aliphatic chains, as well as cholesterol. Major UA force fields are also represented (Berger (Berger et al., 1997), GROMOS-CPK (Piggot et al., 2012), CHARMM-UA (Lee et al., 2014)). In case a user wants to analyze a lipid which is not present in `buildH`, a step-by-step documentation guides the user in the process of creating and supplying his/her lipid description as a json file.

All structure and trajectory read / write operations are handled by the MDAnalysis module (Gowers et al., 2016; Michaud-Agrawal et al., 2011). Mathematical vector operations are performed by Numpy (Harris et al., 2020) and accelerated with Numba (Lam et al., 2015), leading to very decent performances. For example, the reconstruction of all Hs and order parameter calculation on a trajectory of 2500 frames with 128 POPC molecules can be handled in approximately 7 minutes using a single core Xeon processor at 3.60 GHz, whereas it was almost 30 minutes long without Numba.

`buildH` has been implemented with good practices of software development in mind (Jiménez et al., 2017; Taschuk & Wilson, 2017):

- version controlled repository on GitHub https://github.com/patrickfuchs/buildH,
- open-source license (BSD-3-Clause),
- continuous integration through tests,
- and documentation https://buildh.readthedocs.io/.

Some notebooks are provided in the GitHub repository to explain how `buildH` works and how to analyze the data produced. In case of trouble, any user can post an issue on GitHub.

`buildH` is available in the Python Package Index (PyPI) as well as in the Bioconda repository. The current version 1.5.0 of `buildH` is archived in the Zenodo repository (https://zenodo.org/record/5080126) and in the Software Heritage archive (swh:1:dir:eb46a03f6e6188f93bd0b39ab78b4640e777ecd1).

Acknowledgements

The authors thank the community of NMRlipids for useful discussions, especially Samuli Ollila.

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

Antila, H., Buslaev, P., Favela-Rosales, F., Ferreira, T. M., Gushchin, I., Javanainen, M., Kav, B., Madsen, J. J., Melcr, J., Miettinen, M. S., Määttä, J., Nencini, R., Ollila, O. H. S., &

Piggot, T. J. (2019). Headgroup structure and cation binding in phosphatidylserine lipid bilayers. *The Journal of Physical Chemistry B*, *123*(43), 9066–9079. https://doi.org/10.1021/acs.jpcb.9b06091

Bacle, A., Buslaev, P., Fandiño, R. G., Favela-Rosales, F., Ferreira, T. M., Fuchs, P. F. J., Gushchin, I., Javanainen, M., Kiirikki, A. M., Madsen, J. J., Melcr, J., Rodríguez, P. M., Miettinen, M. S., Ollila, O. H. S., Papadopoulos, C. G., Peón, A., Piggot, T. J., Piñeiro, Ángel, & Virtanen, S. I. (2021). https://doi.org/10.5281/zenodo.4748646

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, *91*(1), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E

Berger, O., Edholm, O., & Jähnig, F. (1997). Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal*, *72*(5), 2002–2013. https://doi.org/10.1016/S0006-3495(97)78845-3

Botan, A., Favela-Rosales, F., Fuchs, P. F. J., Javanainen, M., Kanduč, M., Kulig, W., Lamberg, A., Loison, C., Lyubartsev, A., Miettinen, M. S., Monticelli, L., Määttä, J., Ollila, O. H. S., Retegan, M., Róg, T., Santuz, H., & Tynkkynen, J. (2015). Toward atomistic resolution structure of phosphatidylcholine headgroup and glycerol backbone at different ambient conditions. *The Journal of Physical Chemistry B*, *119*(49), 15075–15088. https://doi.org/10.1021/acs.jpcb.5b04878

Catte, A., Girych, M., Javanainen, M., Loison, C., Melcr, J., Miettinen, M. S., Monticelli, L., Määttä, J., Oganesyan, V. S., Ollila, O. H. S., Tynkkynen, J., & Vilov, S. (2016). Molecular electrometer and binding of cations to phospholipid bilayers. *Phys. Chem. Chem. Phys.*, *18*, 32560–32569. https://doi.org/10.1039/C6CP04883H

Douliez, J. P., Léonard, A., & Dufourc, E. J. (1995). Restatement of order parameters in biomembranes: Calculation of c-c bond order parameters from c-d quadrupolar splittings. *Biophysical Journal*, *68*(5), 1727–1739. https://doi.org/10.1016/S0006-3495(95)80350-4

Feller, S. E. (2000). Molecular dynamics simulations of lipid bilayers. *Current Opinion in Colloid & Interface Science*, *5*(3), 217–223. https://doi.org/10.1016/S1359-0294(00)00058-3

Feller, S. E. (2007). Molecular dynamics simulations as a complement to nuclear magnetic resonance and x-ray diffraction measurements. In A. M. Dopico (Ed.), *Methods in membrane lipids* (pp. 89–102). Humana Press. https://doi.org/10.1007/978-1-59745-519-0_7

Gowers, Richard J., Linke, Max, Barnoud, Jonathan, Reddy, Tyler J. E., Melo, Manuel N., Seyler, Sean L., Domański, Jan, Dotson, David L., Buchoux, Sébastien, Kenney, Ian M., & Beckstein, Oliver. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Sebastian Benthall & Scott Rostrup (Eds.), *Proceedings of the 15th Python in Science Conference* (pp. 98–105). https://doi.org/10.25080/Majora-629e541a-00e

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., Capella-Gutierrez, S., Chue Hong, N., Cook, M., Corpas, M., Flannery, M., Garcia, L., Gelpí, J. Ll., Gladman, S., Goble, C., González Ferreiro, M., Gonzalez-Beltran, A., Griffin, P. C., Grüning, B., … Crouch, S. (2017). Four simple recommendations to encourage best practices in research software. *F1000Research*, *6*, 876. https://doi.org/10.12688/f1000research.11407.1

Kiirikki, A., & Ollila, S. (2021). *NMRlipids databank*. https://github.com/NMRLipids/Databank

Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D., & Pastor, R. W. (2010). Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *The Journal of Physical Chemistry B*, *114*(23), 7830–7843. https://doi.org/10.1021/jp101759q

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based python JIT compiler. *LLVM '15: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. https://doi.org/10.1145/2833157.2833162

Lee, S., Tran, A., Allsopp, M., Lim, J. B., Hénin, J., & Klauda, J. B. (2014). CHARMM36 united atom chain model for lipids and surfactants. *The Journal of Physical Chemistry B*, *118*(2), 547–556. https://doi.org/10.1021/jp410344g

Lyubartsev, A. P., & Rabinovich, A. L. (2011). Recent development in computer simulations of lipid bilayers. *Soft Matter*, *7*, 25–39. https://doi.org/10.1039/C0SM00457J

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., & Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, *32*(10), 2319–2327. https://doi.org/10.1002/jcc.21787

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(33). https://doi.org/10.1186/1758-2946-3-33

Ollila, S. (2014). *On the signs of the order parameters*. http://nmrlipids.blogspot.com/2014/04/on-signs-of-order-parameters.html

Piggot, T. J., Allison, J. R., Sessions, R. B., & Essex, J. W. (2017). On the calculation of acyl chain order parameters from lipid simulations. *Journal of Chemical Theory and Computation*, *13*(11), 5683–5696. https://doi.org/10.1021/acs.jctc.7b00643

Piggot, T. J., Piñeiro, Á., & Khalid, S. (2012). Molecular dynamics simulations of phosphatidylcholine membranes: A comparative force field study. *Journal of Chemical Theory and Computation*, *8*(11), 4593–4609. https://doi.org/10.1021/ct3003157

Taschuk, M., & Wilson, G. (2017). Ten simple rules for making research software more robust. *PLOS Computational Biology*, *13*(4), e1005412. https://doi.org/10.1371/journal.pcbi.1005412

Tieleman, D. P., Marrink, S. J., & Berendsen, H. J. C. (1997). A computer perspective of membranes: Molecular dynamics studies of lipid bilayer systems. *Biochimica Et Biophysica Acta (BBA) - Reviews on Biomembranes*, *1331*(3), 235–270. https://doi.org/10.1016/S0304-4157(97)00008-7