NMRLipids Project on Polarizable Lipid Force Fields
Data Contribution Guidelines

Tuesday 31st March, 2020

In the NMRLipids Project VI on polarizable lipid force fields, we are aiming at giving a comprehensive comparison of the performance of the currently available polarizable models for the lipid bilayers. Our main experimental observables are the NMR deuterium order parameters and the effective correlation times.

The general description and the goals of this particular project are stated in this blog post (a link here, when we have the post online). For more information on the overall NMRLipids project, please read here.

1   Details of the Codes Employed in this Project

In order to automatize the analysis and reduce the human effort, we have compiled a Python script. This script, "*AddData.ipynb*", is a Python script created using the Jupyter notebooks and available at the "scripts" directory of the GitHub repository for this project. This is our main script that reads the trajectories from any public repository using the unique Digital Object Identifier (DOI), does necessary preprocessing of the trajectories, calls other scripts (*OrderParameter.py*, *corr_ftios_ind.sh*, and *corrtimes.py*) to calculate the deuterium order parameters and effective correlation times, and writes the output to a directory with an unique name which is derived from the DOI of the data set.

AddData.ipynb script requires the simulation description part to be supplied from the user. This part is located in the beginning of the script, where the user needs to define the DOI for the simulation repository, a definition file that contains the necessary atom indexing to calculate the order parameters, employed software, force field, force field source, and the names of the trajectory files, and finally a working directory where the files will be written. The rest of the calculation process is automatically done and the user will end up with a directory that contains a README file plus files that contain the order parameters and the effective correlation times. Please note that, all these files will be automatically generated and the directory that contains them will have a unique name derived from the supplied DOI number. Therefore, no change/alterations to the files or the file names will be necessary.

As of Tuesday 31st March, 2020, *AddData.ipynb* is capable of processing and analyzing the trajectories of *xtc* (Gromacs) and *dcd* (NAMD) formats. In case your trajectories are in a different file format, please either make the necessary changes to the *AddData.ipynb*and push it back to the GitHub repository, or contact us so that we can handle your request and update the script according

to your data format.

For *AddData.ipynb* to work, it needs two additional Python libraries. Both MDTraj and MDAnalysis are required to be installed on your computer. MDAnalysis is required by the *OrderParameter.py*. MDTraj is required to convert trajectories, in case they differ from Gromacs' *xtc* format, to *xtc* format so that the *AddData.ipynb* can function.

In addition to the MDAnalysis and MDTraj, the you need to have the scripts *OrderParameter.py*, *corr_ftios_ind.sh*, and *corrtimes.py* in your working directory. These scripts are also available at our GitHub repository. An additional file, which we refer as the description file, is also required to calculate the order parameters and the effective correlation times. The description is essentially a mapping file that matches the order parameter definitions to the atom names for a given force field. An example of such file is available here.

The trajectories used for the analysis must be on a publicly available server that assigns a unique DOI to the data set. One example of such server is Zenodo. We kindly ask all contributors to upload their data to Zenodo under the NMRLipids community. The data description required by the Zenodo is not compulsory for our purposes, yet a proper documentation of the data sets will surely be beneficial for all of us. We encourage potential contributors to take a look at the data sets available under the NMRLipids tag to get an idea about the documentation of the data sets.

2 Data Contribution Guideline

If you want to contribute to the project with your data, please follow these steps:

1. Upload data set to Zenodo and obtain the unique DOI number for your data set,

2. Fork NMRLipidsVIpolarizableFFs repository at GitHub to your own GitHub account,

3. Download the scripts *AddData.ipynb*, *OrderParameter.py*, *corr_ftios_ind.sh*, and *corrtimes.py* to your working directory,

4. Either create or download a description file specific to the force field you used,

5. Edit the "Simulation description file" section of the *AddData.ipynb* to match your info,

6. Run the *AddData.ipynb* script to calculate the order parameters and effective correlation times,

7. Push the directory (which should have name like "10.5281:zenodo.3557459.0") to the NMRlipidsVIpolarizableFFs/Data/Simulations directory on GitHub,

8. Comment on the NMRLipids blog about your data contribution.

In case you do not want to run the analysis on your own, please let us know the whereabouts of your raw data so that we can run the analysis for you.

If you do not have any data at the moment but are willing to run simulations to contribute, please leave a comment on our blog so that everyone knows which specific trajectories you will be

supplying. This is important for all of us to keep track of the upcoming simulations and identify the missing parameter sets.

Please note that, we will need to review your data contribution before it appears in the GitHub repository. Then, we will merge your request to the main repository or will contact you if there is anything missing or unclear.

An important note: Please do not change the name of the directory that the *AddData.ipynb* creates while uploading your data to our GitHub repository. This is very important for indexing the data contributions.

Another Important note: The analysis scripts can change within time as we add or remove some code. Therefore, please make sure that you have updated your GitHub fork and scripts to the most recent versions.

3   Authorship and Data Ownership

As the data contributor, you have your rights on your data set, according to the licensing option you choose on the public server where you upload your data. By submitting your data set to the NMRLipids community on Zenodo, you are agreeing that your data will be used in the NMRLipids projects. Please read the licensing options on Zenodo (or your favorite public data server). By default and as the community we are encouraging everyone to use Creative Commons licensing scheme (whichever version suits you) and make your data set publicly available to whomever needs it. Please note that, the data server you are using for uploading your data set generates a DOI.

As the NMRLipids community, we believe the science should be practiced openly and whomever wants to contribute should be welcomed. Therefore, by submitting your data you will become a member of the NMRLipids community. Any contribution, regardless in form of contributing a data set or commenting on the ongoing discussion in our blog, you **will** be offered a joint authorship in the upcoming NMRLipids publications. Of course, it is your prerogative to accept or decline the authorship. We will do our best to publish the results from the project in an open-access journal and will inform everyone about its progress through our blog and email list.

Specific to the NMRLipids VI project on polarizable models, in case there is a publication, Batuhan Kav of Forschungszentrum Jülich, Germany, will be the corresponding author. We will offer everyone who contributed to the project a joint authorship. Should you accept our invitation, your name will be included in the publication in the alphabetical order. Regardless of the journal's publishing policy, we will inform everyone about the referee comments, our response to the referees, etc. and will ask for your contribution.

In case you have any questions or concerns, please do not hesitate to leave a comment on our blog, GitHub repository, or send us an email.

Please do not forget that this is an open collaboration that can only survive and thrive with your contributions. Any contribution, regardless it is a comment on the blog, adding a few lines to the analysis scripts, or just supplying raw data is invaluable to us.