Instructions

AddData.py

AddData.py is a script that builds a database that contains a dictionary file and analysis data of each simulation. The dictionary file contains information about the simulation. The script also calculates order parameters of all CH bonds of the lipids in the simulation. To add a simulation it must be first uploaded to Zenodo (www.zenodo.org). The trajectory and topology files of the simulation are downloaded to the working directory from Zenodo but these are not saved into the database. To add a simulation to the database the user has to give some essential information about the simulation. This is done by writing a info file (*INFO.py) which is passed to AddData.py.

AddData.py requires GROMACS and MDAnalysis library to be installed.

Run AddData.py as follows:

python3 AddData.py exampleINFO.py

There is also a script runAddData.sh that can be used to loop over several info files to add many simulations at one go.

Simulation dictionary

A simulation dictionary contains information about the simulation. Some of the information is provided by the user. The numbers of lipid molecules, solvent and ions are automatically read from the files and so are the simulation temperature and trajectory length. All this information is saved to a file named README.yaml.

User input

DOI

DOI is the DOI access number of the simulation. Use the version DOI.

SYSTEM

SYSTEM is name of the system.

MAPPING

MAPPING provides the name of the mapping file for a lipid. It must be given as a string of alternating lipid name and the corresponding mapping file name separated by a comma. The existing mapping files are in the directory named "mapping_files". If a mapping file of a certain lipid does not exist the user must construct it.

The purpose of a mapping file is to circumvent the problem caused by different atom naming conventions used by different force fields. The first column of a mapping file contains general atom names. The second column contains the name of the atom as it is in the force field. If the lipid consists of several residues which is the case with some AMBER force fields, then a third column is needed which contains the name of the residue to which each atom belongs to.

SOFTWARE

SOFTWARE stands for the name of the software used for running the simulation. The options are GROMACS, AMBER, NAMD, CHARMM and OPENMM. So far, only simulations run with GROMACS are accepted by the script.

FF

FF is the name of the forcefield used in the simulation.

FF_SOURCE

FF_SOURCE tells where the forcefield parameters are taken from.

FF_DATE

FF_DATE is the date when the forcefield parameters were created. The format is day/month/year.

TRJ

TRJ stands for the name of the trajectory file.

TPR

TPR stands for the name of the topology file.

PREEQTIME

PREEQTIME means the time used for pre-equilibrating the system. This should be in nano seconds.

TIMELEFTOUT

TIMELEFTOUT stands for the length of the simulation in nano seconds to be left out at the beginning of the simulation.

UNITEDATOM

The handling of united atom simulations is enabled by a separate script called buildH_calcOP.py. In case of a united atom simulation, the user has to give the names of the lipids and the corresponding names of those lipids as they are in the dic_lipids.py dictionary used by buildH_calcOP.py script. In case of an all atom simulation this can be given as an empty string.

Molecule names

The user must also provide the names of the molecules, ions and solvent that are used. These are obtained from the tpr file. If a lipid consists of more than one residue the name of the head group residue is given.

dir_wrk

Finally the user has to give the path of the working directory.