Analysis of GROMACS 2020 simulations with Jupyter notebooks

Introduction

We have prepared jupyter notebooks which allow a more straightforward analysis of GROMACS simulations. Currently the notebooks are based on running and collecting the output from GROMACS programs, but we also use the MDAnalysis toolkit for some analyses. To do this you need to install the python software on your own PC, copy the data down from M100 and run jupyter locally.

Requirements

1. python3
2. web browser
3. notebook file and library (`EX4Cutils.py`)
4. GROMACS 2020.x or higher
5. GROMACS xtc, tpr, pdb and index files of a simulation.

Setup

You need native Linux or Linux running under WSL (Windows). Gromacs can be installed in Ubuntu via:

```
sudo apt update
sudo apt install gromacs
```

Preparation of GROMACS trajectory

Make sure PBC and solvent have been removed. In many cases the following works:

```
gmx trjconv -f traj.xtc -pbc nojump -center -o traj-nopbc.xtc -s topol.tpr
```

When requested, use "Protein" for centering and output. If the trajectory is large then consider the `skip` parameter:

```
gmx trjconv -f traj.xtc -skip 10 pbc nojump -center -o traj-nopbc.xtc -s topol.tpr
```

A pdb file is best generated from the first frame of this trajectory.

```
gmx trjconv -f traj-nopbc.xtc -dump 0 -o traj.pdb
```

This version of the software also requires an index file.

```
gmx make_ndx -f topol.tpr -o index.ndx
```

Select 'q' to save and exit.

Virtual environment

Make sure you have python3 and GROMACS 2020 installed. We recommend a virtual environment (although not essential):

```
python -m venv analysis
source analysis/bin/activate
```

Library install

Warning: this may take some time, but only needs to be done once.

```
pip3 install matplotlib pandas
pip3 install MDAnalysis
pip3 install numpy
pip3 install notebook
```

If you have difficulty installing Pandas then try:

```
pip3 install ipykernel --upgrade
python3 -m ipykernel install --user
```

Difficulties with matplotlib can be resolved by loading cython:

```
pip install cython
pip install matplotlib
```

For visualization with nglview (optional, not currently working on M100):

```
pip3 install nglview
pip3 install ipywidgets
jupyter nbextension enable --py widgetsnbextension
```

If not using a virtualenv you may need to the last command with sudo.

Analysis

Running the analysis notebook

First make sure you that you have the `EX4Cutils.py` library file in the same directory from where you run jupyter. Then launch:

```
jupyter notebook
```

Workflow

1. A browser should open automatically, otherwise copy and paste the link provided into an open one.
2. Within the browser locate and open the `single-analysis.ipynb` file.
3. Go to the section "File definition" and define the locations of the files.
4. Start from the top of the notebook and allow the libraries to be loaded
5. You can then follow the procedure to perform the analysis: note that with the exception of the RMSD and radius of gyration, all the analyses require GROMACS. Note that the clustering analyses may take some time.

Hints

Hints for using jupyter notebooks: - cell types can be python code or markdown - this can be changed from the menu -inserts a carriage return, while executes the cell if containing python code or renders the cell content, if containing markdown - cells containing % will run unix commands in the shell - output is plotted in line with matplotlib

Saving the analysis

Apart from the notebook it is also possible to save the analysis in other ways: - pdf files (easiest way is via the browser) - html

```
jupyter nbconvert --execute --to html notebook.ipynb
```