NMRlipids online meeting: Dabank and polarizable force fields

Date: 26.2.2021

9.00-10.00 Samuli Ollila: Welcome and current status of the NMRlipids databank

10.00-11.00 Batuhan Kav: Current status of the NMRlipics VI on the polarizable force fields

11.00-12.00 Break

12.00-13.00 Discussion on NMRlipids databank

13.00-14.00 Discussion on NMRlipids VI project

14.00-14.30 General discussion and closing of the meeting.

Quality evaluated atomistic resolution MD simulations of biologically relevant lipid mixtures in NMRlipids databank

	PC	PE	PG	PS	chol	CL	GM1	PIP	PA	DAG
PC	X	X	X	X	X					
PE		X								
PG			X							
PS				X						
chol										
CL										
GM1										
PIP										
PA										
DAG										

NMRlipids databank
general properties

- Overlay databank: *NMRlipids databank contains indexed links to the data. The actual MD simulation data is currently in Zenodo, but could be in any stable location.*

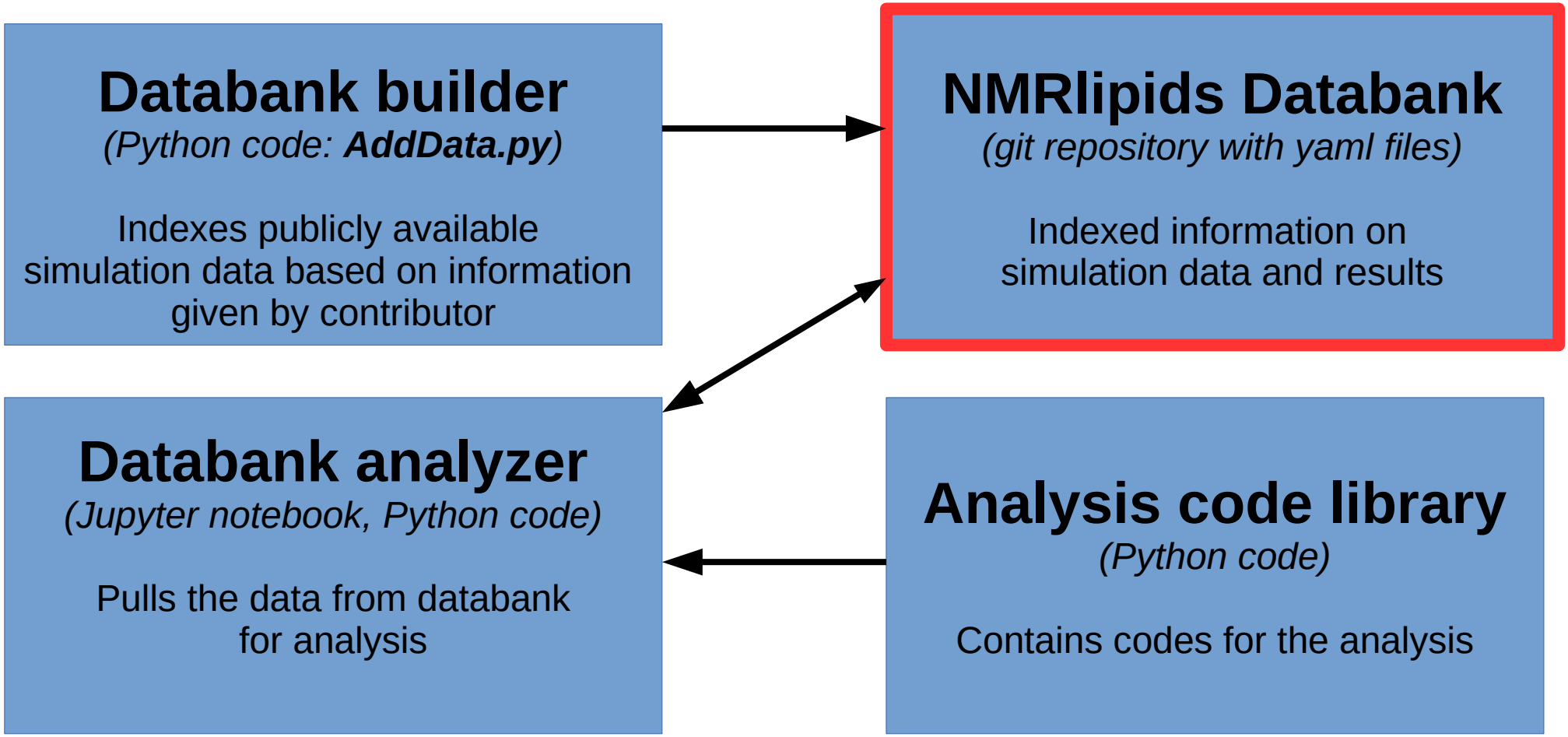
- Analysis of the data: *NMRlipids databank enables flexible analysis of the content.*

- Quality evaluation: *NMRlipids databank contains a quality evaluation protocol that is applied to all contributed datasets. Also the quality evaluation results are also stored in the databank.*

NMRlipids databank
expected applications

- Force field evaluation: *What is the best force field for my application?*
- Reference simulations: *For example, reference pure bilayer simulations for membrane-protein interaction studies.*
- Analysis of bilayer properties from large datasets: *For example, calculate P-N vector angle from all available PC and PG simulations.*
- Exercise and example for sharing simulation data: *“PDB” for simulations?*

NMRlipids databank structure



Databank builder: AddData.py

https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/DataBankINFO/AddData.py

- **Compulsory input required from the contributor**
  - **DOI** (currently works only for Zenodo)
  - **Name of mapping located in**
    https://github.com/NMRLipids/NMRlipidsIVPEandPG/tree/master/scripts/mapping_files
  - **Software** (AMBER, CHARMM, OPENMM, but works currently only for GROMACS)
  - **Name of trajectory file**
  - **Name of topology file**
  - **Pre-equlibration time** = time simulated before uploaded trajectory
  - **Discarded equilibration time** = time that should be discarded from the beginning of the trajectory
  - **United atom information** (empty string for all atom simulations)
  - **Molecule names** = Molecule names in the simulation corresponding standard names in Databank, e.g., POPC, SOL
  - **Working directory** = Directory in local computer which is used to build the databank information.
- **Optional input from the contributor**
  - **System description** = Free text description of the system
  - **Force field** = Name of the force field used in simulation
  - **Force field source** = Origin of force field parameters, e.g., CHARMM-GUI
  - **Force field date** = Date when force field was obtained
  - **Names of force fields for individual molecule names**. For example, FFNA = ECCNa

Databank builder: AddData.py

https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/DataBankINFO/AddData.py

- **Automatically extracted information by AddData.py**

  - **Number of molecules:** Number of molecules (e.g., NPOPC, NSOL, NSOD, etc), lipids separately for leaflets.

  - **Temperature**

  - **Length of the trajectory**

- **Automatic analysis by AddData.py**

  - **Order parameters of all C-H bonds**

- **Output of AddData.py**

  - **All information stored in README.yaml files**

  - **Order parameters stored in files**

  - **These are stored in the folders based hash IDs. Beta version:**

  https://github.com/NMRLipids/NMRlipidsIVPEandPG/tree/master/Data/Simulations

- **Usage of AddData.py**

  - **python3 AddData.py *SimulationInfoFile.py***

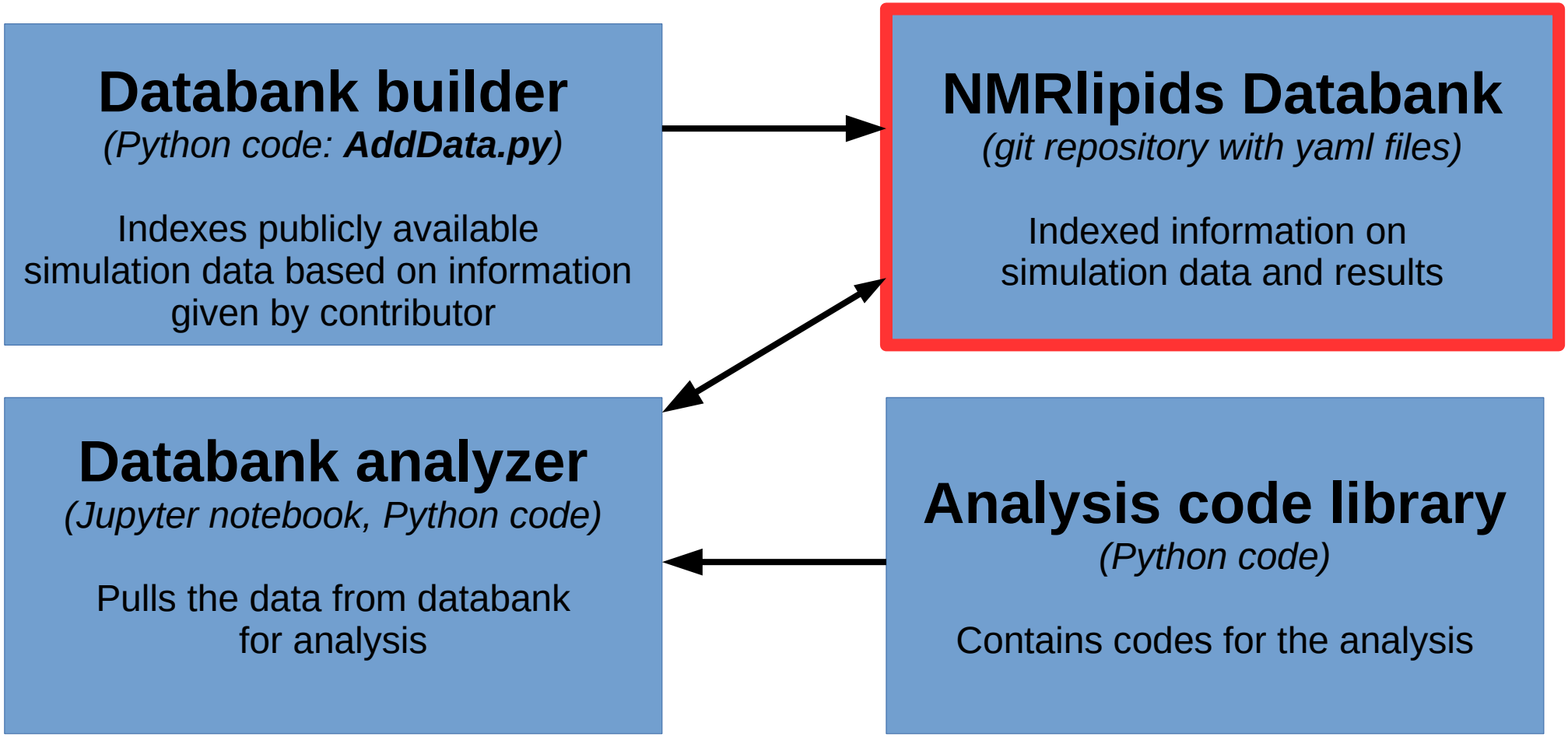
- - ***SimulationInfoFile.py* currently available at:**
  https://github.com/NMRLipids/NMRlipidsIVPEandPG/tree/master/scripts/DataBankINFO

NMRlipids Databank

- Beta version:

  https://github.com/NMRLipids/NMRlipidsIVPEandPG/tree/master/Data/Simulations
- Each folder corresponds one simulation
- Folders are named according to the hash of trajectory and tpr file
- **Folders contain README.yaml which should contain all the relevant information on the simulation! Is something missing?**
- Currently folders contain also automatically calculated order parameters (maybe quality evaluation in the future)

NMRlipids databank structure



Databank analyzer

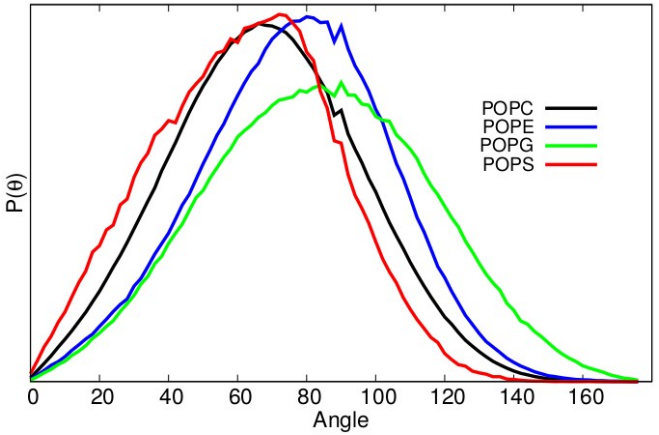
- Goes through the README.yaml files in the databank
- Performs wanted analysis for selected simulations
- Results can be saved in separate results databank with the same indexing
- Result databanks can be browsed in similar manner for plotting

Databank Analyzer Examples

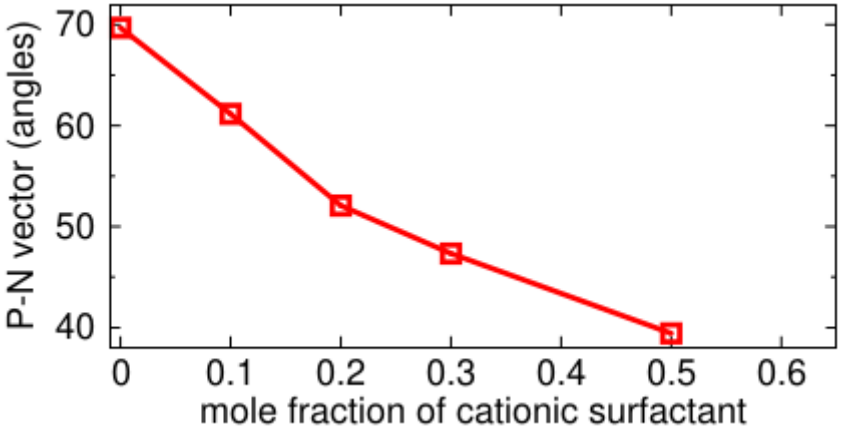
1) Calculate the P-N vector angles of POPS, POPE, POPG and POPC lipids from each simulation

Code: https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/calcPNvectors.py

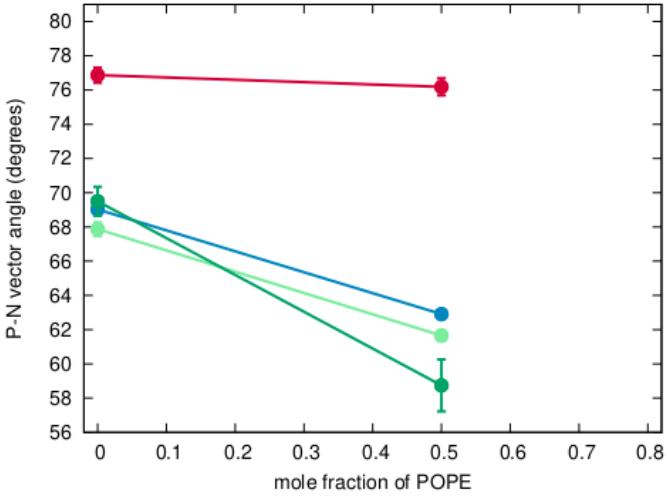
Results: https://github.com/NMRLipids/NMRlipidsIVPEandPG/tree/master/Data/HGorientation

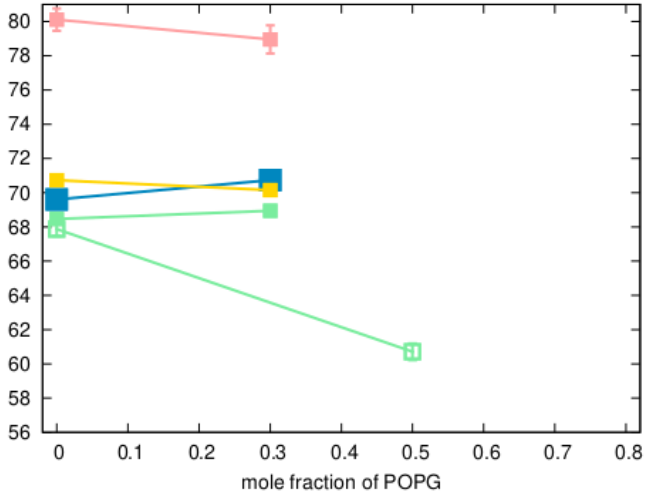


POPC

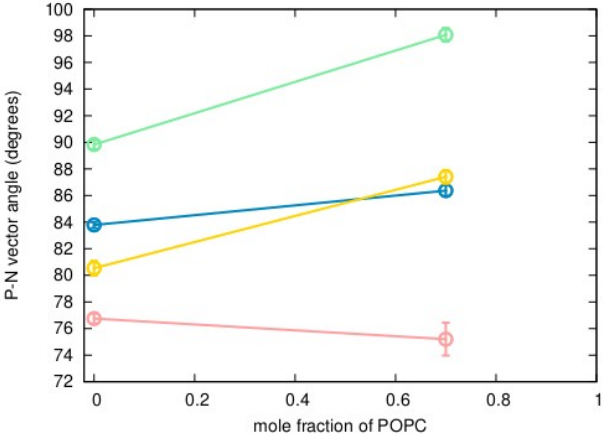


POPC





POPG



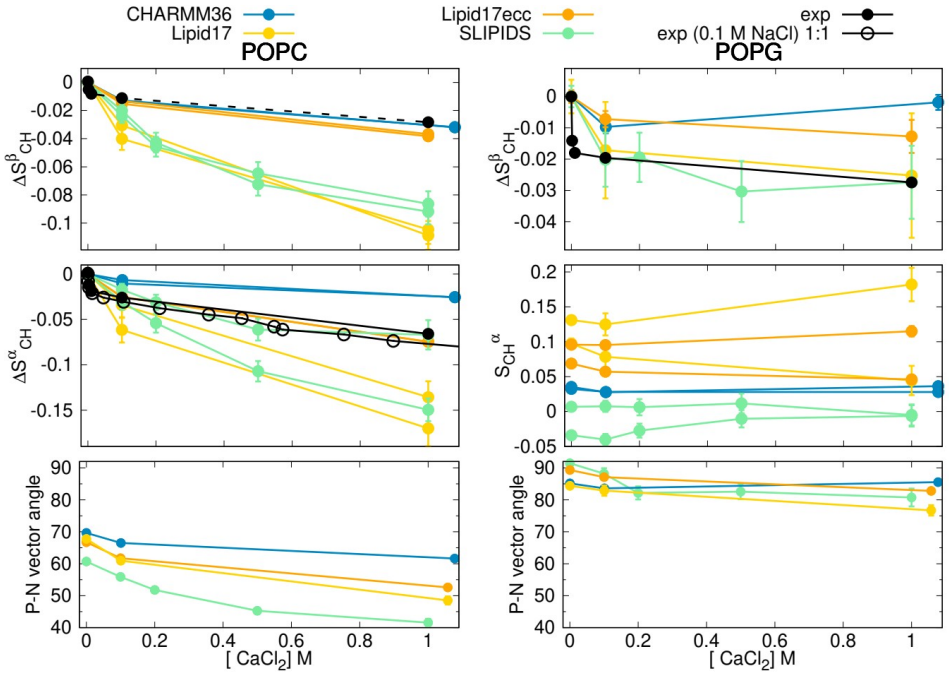
Databank Analyzer Examples

2) Find data for order parameter changes upon addition of CaCl_2 from all available POPC:POPG mixtures

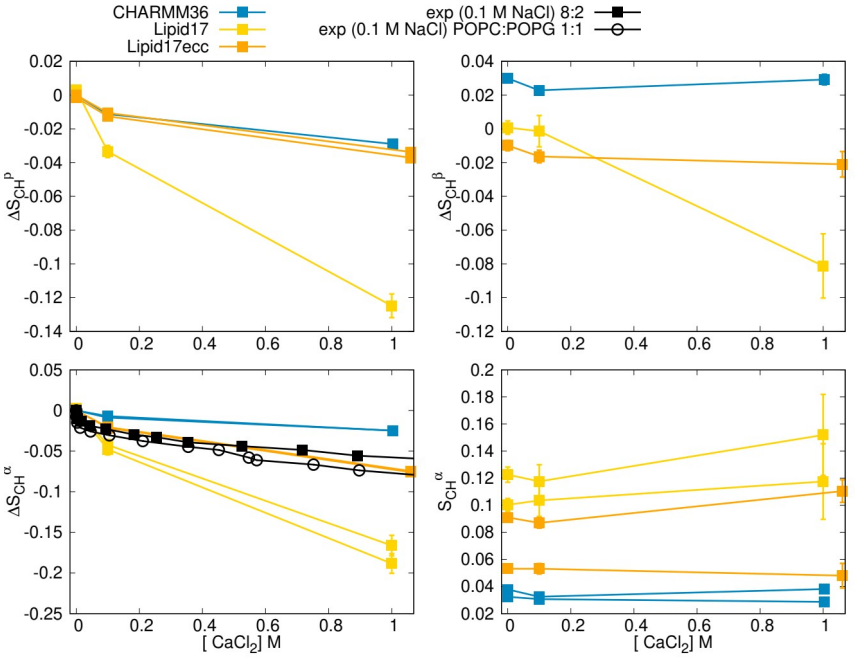
Code: https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/plotOPsWITHsalt.ipynb

Results:

POPC:POPG (1:1)



POPC:POPG (4:1)



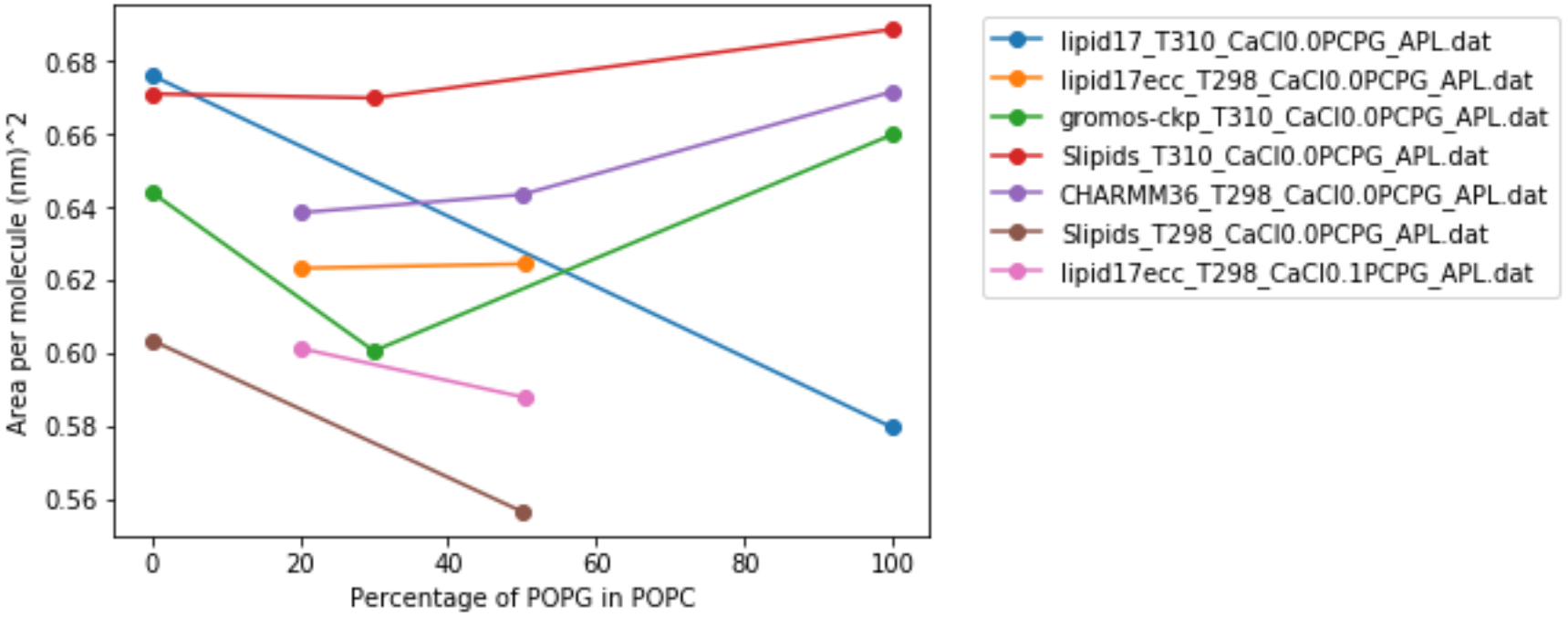
Databank Analyzer Examples

3) How area per lipid changes in PC:PG lipid mixtures as a function of PG concentration

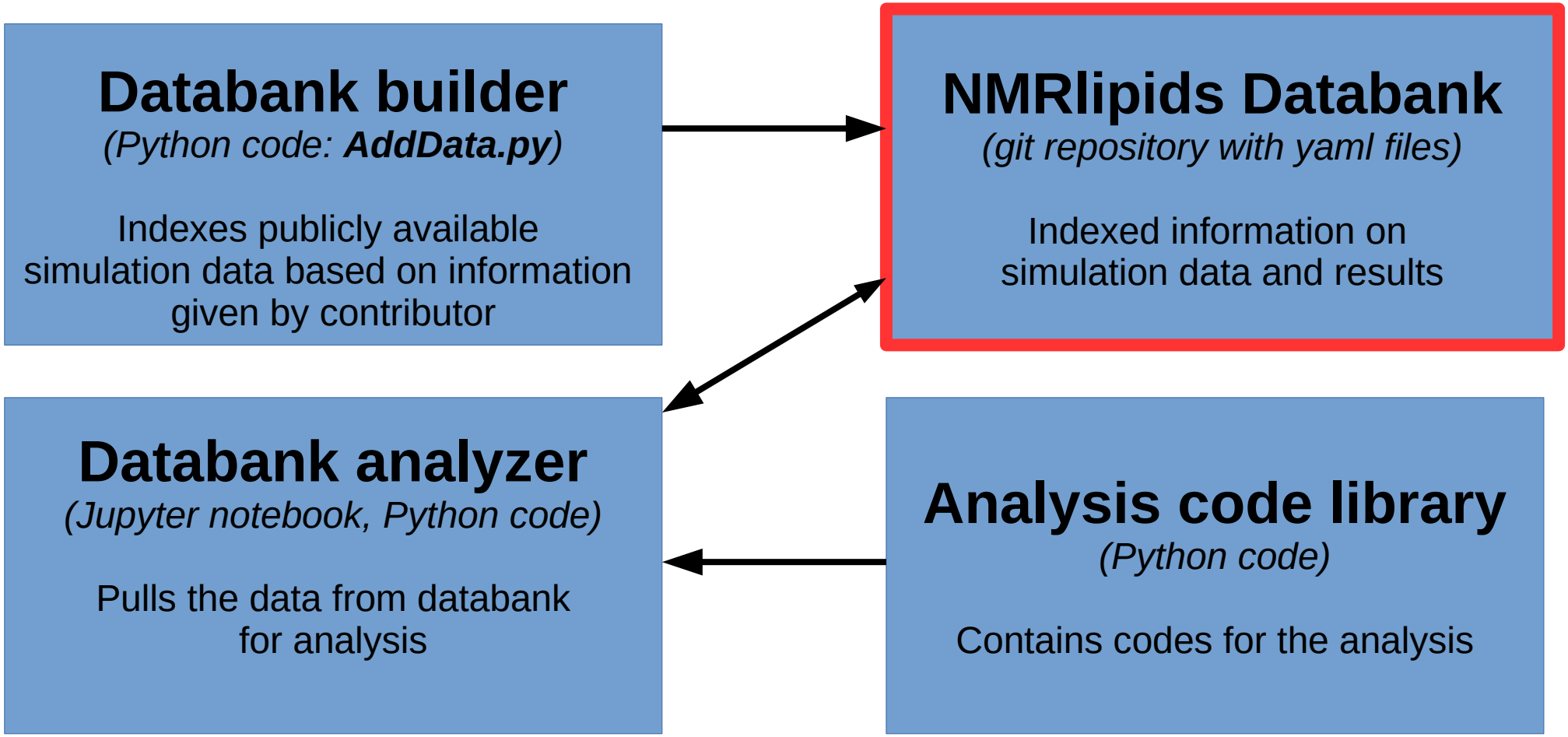
Code: https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/calcAPL.py

https://github.com/NMRLipids/NMRlipidsIVPEandPG/blob/master/scripts/plotAPLs.ipynb

Results:



NMRlipids databank structure



NMRlipids databank publication plan

- **Article describing the databank and highlight applications will be prepared.**
- **At least all trajectories contributed to the NMRlipids will be included (approximately 300-400 trajectories currently).**
- **Possible highlight applications:**

  - Quality ranking of all simulation

  - Analysis of rare phenomena using large datasets, such as water permeation through bilayers or lipid flip-flops

  - Example of analysis useful for community who are typically not using MD simulations, such as T_1 spin relaxation times of water near membranes that are used in MRI imaging

- **Databank will be located in a GitHub repository.** **Under which kind of licence?**
- **NMRlipids authorship rules will be applied in the first publication of the databank** (authorship will be offered to all contributors and order is alphabetical) **with two exceptions: Samuli Ollila will the last author and Anne Kiirikki will be the first.**
- **The authorship policy in the future publications regarding NMRlipids databank have to be carefully discussed in the near future.**

NMRlipids databank publication plan

- Needs to be done for publication:
  - **Experimental data to the databank and automatic quality ranking of all simulations** (in progress by Anne Kiirikki)
  - **Incorporation of all data into the databank.** To be done when content of REAME.yaml files agreed.
  - **Analysis of highlight applications.**

Issues to be discussed

- **What should we store in README.yaml files? Is there something that should be added?**
- **What should be analyzed automatically? Currently only order parameters.**
- **Should we extend to other than Gromacs simulations before first publication?**
- **License of the databank GitHub repo?**
- **Authorship in future NMRlipids databank publications.**