13-11-2024 Code modification:

Trajectory class:

* Logging is now done via logging module, rather than print and saving. output\_directory and output\_prefix needs to be moved to \_\_init\_\_ function since we would like to log into the file starting from the beginning. However, if you rerun only the run function, the log will be appended to the existing file.
* Added two arguments for Trajectory class:   
  1. hetatm\_resid: int, if supplied the record type of this specific resid will be converted to HETATM.  
  2. custom\_terminal\_oxygens: list, the terminal oxygen atom names the user provided will be modified to O and OXT (compatible with PROPKA).
* 3. custom\_resname\_correction: dict, if resname in our predefined dictionary does not cover the user’s topology, even if it is an amino acid residues, user can use this argument to add the correction.
* Swapped the order of argument topology\_file and trajectory\_file.
* Input check has been optimised to check HETATM, terminal oxygens and resname.
* Formatted the docstring in Trajectory class and run.
* Default chain = ‘A’ in the run function.
* sort\_pka\_function simplified.
* disulphide\_bond\_detection removed the print statement. The listed cysteines that is forming disulphide bond will be saved in the output instead.
* pka\_iterator optimized. Originally the pKa and buriedness data for each frame is appended to CSV. Now everything will be saved in memory and only written in sort\_pka\_function.
* Tried different methods but I still couldn’t redirect the warning from propka (stdout) from different workers to a file.
* The corresponding residue of mutation will now be logged.

Clustering:

* Stuart noticed a bug in \_determine\_charge\_center\_. Primarily due to the selection = universe.residues[residue\_index - 1]. If the resid does not start with 1, this would become an error. Changed it to selection = universe.select\_atoms(f‘resid {resid}’)
* Swapped order of topology\_file and trajectory\_file to be consistent with MDAnalysis.
* Why “k-medoids++” but not “heuristic”? IM tried heuristic, but warning occur where it shows cluster X is empty:

A black screen with blue and white text

Description automatically generated

https://stackoverflow.com/questions/62215324/sklearn-kmedoids-returns-empty-clusters

* I can’t be bothered to change the writing log to use the logging module.

Visualisation:

* Removed the NtermCap\_atom\_name and CtermCap\_atom\_name in argument. Now it can be guessed internally.
* Chain argument (default = ‘A’) added so user need to provide which ensure the correct chain is labelled (if the topology it provides contain two chains with the same resid, this would cause a bug.)
* Output\_directory, output\_prefix
* Correlation\_threshold added so that only residues with abosolute correlation coefficient above this threshold will be saved in the PyMOL session.
* Each residue is now an object in the PyMOL session (rather than selection), so user can hide it by clicking the object.

Github:

* Created a branch for AF once everything is updated
* Test installation added to README.md
* Run time added to README.md
* Removed unnecessary files through .gitignore

Tutorial:

* All variables and import put at the top.
* Argument descriptions deleted in the notebook, but some arguments are explained in the comment next to the arguments
* Images resized
* Print which trajectory (+mutation) is being processed at the moment
* Plot title replaced to actual prefix of trajectory.
* Did not add progress bar since it is not clean.
* Markdown changed.