Contribution ID: **30** Type: **not specified**

Physically-oriented clustering of conformations in multi-molecule systems

Unsupervised machine learning techniques are widely used to analyze the extensive data generated by molecular modeling. In particular, some tools have been developed to cluster configurations from classical simulations with a standard focus on individual units, ranging from small molecules to complex proteins. Since the standard approach computes the root-mean-square deviation (RMSD) of atomic positions, accounting for atom permutations is crucial to optimizing clustering of multi-molecule systems featuring identical molecules. To address this issue, we developed the clusttraj program, a solvent-informed clustering package that corrects inflated RMSD values by finding optimal pairings between configurations. The program combines reordering schemes with the Kabsch algorithm to minimize the RMSD of molecular configurations before running a hierarchical clustering protocol. By considering evaluation metrics, one can automatically determine the optimal threshold and compare available linkage schemes. The program's capabilities will be illustrated by considering various systems, ranging from pure water clusters to solvated proteins and oligomer chains in different solvents. Ultimately, we reduce the data complexity and obtain a subset of conformations whose dependence on cluster-related parameters and representativeness with respect to desired properties will be discussed. clusttraj is implemented as a Python library and can be used to cluster generic ensembles of molecular configurations beyond solute–solvent systems.

Author: RIBEIRO, Rafael (Universidade de São Paulo)

Presenter: RIBEIRO, Rafael (Universidade de São Paulo)