Solvent-Informed Clustering of Molecular Mechanics Trajectories

Rafael Bicudo Ribeiro[1]*, Márcio T. do N. Varella[1], Henrique Musseli Cezar[2]
[1]*University of São Paulo, São Paulo, Brazil*
[2]*Hylleraas Center for Quantum Molecular Sciences, Oslo, Norway*

*email: rafael.bicudo.ribeiro@usp.br

Classical mechanics simulations are widely applied from biomolecular systems to materials science and can produce a very large amount of data. Several approaches have been proposed to track the time evolution of molecular conformations [1-3], but generally restricted to isolated molecules. To account for surrounding effects, we employed a machine learning clustering algorithm that includes solvent molecules in the Root Mean Square Deviation (RMSD) calculation. We compared different linkage methods inside the hierarchical clustering framework for three representative solvent molecules, *i.e.*, water, chloroform and toluene. The amount of solvent considered in the configurations and the similarity cutoff proved to be relevant factors. Increasing the number of solvent molecules raises the overall deviation, while changing the cutoff, not only impacts the size of the clusters, but also the quality of the procedure. However, by computing the silhouette coefficient we were able to determine the optimal cutoff for each system and method. As a result, we establish a quantitative approach to reduce the amount of data and therefore, study bulk properties from a compact and representative set of configurations.

References

1. *J. Chem. Inf. Model.* **58**, 2178-2182 (2018)
2. *Bioinformatics*, **38**(1), 73–79 (2022).
3. *J. Chem. Theory Comput.* **9**, 3084−3095 (2013)