Cluster analysis of AcrB protein molecular dynamics conformations

Núbia Souza Prates[1], Adriano V. Werhli[2], Karina dos Santos Machado[2]

[1]*Biological Sciences Institute - Universidade Federal de Minas Gerais*
[2]*Center of Computacional Sciences - Universidade Federal do Rio Grande*

Molecular dynamics (MD) simulation is a computational method widely used to study biological macromolecules at an atomic level. The classical MD simulation is based on numerical solution of Newton's equations of motion. Typically this method may generate as result a large amount of data to be analyzed. These data consists in thousand of conformations that can be used, for instance, to understand the macromolecule flexibility and/or to incorporate this flexibility in docking studies. Due to the complexity of these data, cluster analysis is one approach that can be applied to MD trajectories to reduce this data, identifying representatives conformations of each cluster. First we performed a 50 ns MD simulation of AcrB efflux pump protein using GROMACS package. Protein conformations were taken from the 50 ns MD trajectory with intervals of 50, 20 and 10 ps totaling 3 datasets of conformations having 1,000, 2,500 and 5,000 protein structures each. Thus, in this work we propose to compare three different clustering algorithms for MD simulations considering as input 3 datasets of protein conformations. We applied the clustering algorithms Gromos, Single-Linkage and Jarvis-Patrick implemented in GROMACS. Using the Root-mean-squared deviation (RMSD) as similarity measure with a cutoff from 0.1 to 0.5 nm. Increasing the RMSD values we notice that for Single Linkage and Jarvis-Patrick the number of clusters decreases and all structures tends to remain in a unique cluster. The same was observed for Gromos, where higher values of RMSD implies in all the conformations in the same cluster or leading to the formation of fewer overpopulated clusters. The amount of clusters composed by a single structures was higher in Jarvis-Patrick clustering results compared with the others algorithms using the cutoff 0.17 nm. In the implementation of the clustering algorithm in GROMACS there is no internal or external cluster validation indices to evaluate which algorithm presented the best clustering results. Thus, we consider as best clustering results those clusters with more than 2 structures, fewer singletons clusters or those clusters with approximately the same distribution of protein conformations. Compared to others algorithms Gromos produces significant clustering results, presenting the same behavior with all datasets. We also evaluate the Free Energy of Binding (FEB) by molecular docking experiments considering as receptor representative conformation of each cluster and a ligand called NLM02. Additionally, Gromos presented the best results of docking experiments with a minimum FEB of -9,8 kcal/mol, while -9,4 for Single-Linkage and Jarvis-Patrick.

Supported by: CAPES, CNPq.