dQCuts: Simplifying Analysis of Large Molecular Dynamic Trajectories Using Higher Order Statistical Signatures

Vacaliuc, Gabe

Analysis of protein chains has shifted into a computational biology realm in recent years because of insufficient experimental imaging. Proteins are modeled using molecular dynamics (MD) and a simulation is played out over several microseconds. However, because proteins move faster than nanosecond time scales, millions of frames of simulation are created. Thus, the development of analytic software is needed. I present dQCuts, a new software package which automatically clusters trajectories of protein conformations in order to isolate distinct states during protein folding. dQCuts is the culmination of multivariate analysis and spectral clustering tailored for MD simulation analysis. Using a test simulation consisting of 123 trajectories modeling the protein Ubiquitin, dQCuts managed to isolate 6 distinct folding substates. Knowledge of these states allows researchers to reference structures encountered in neutron scattering experiment, as well as potentially finding a binding site to be used in drug discovery.