Application of Principal Components Analysis in Protein Molecular Dynamic Simulations

Yinling Liu
Sep. 13th, 4:00PM, Hunter 100

Principal Components Analysis (PCA) is a common dimension reduction technique for identifying statistical patterns in data of high dimension. This literature seminar is designed to give you an understanding how PCA can be implemented in the exploration and study of the biomolecular structure and dynamic simulations. In molecular dynamic (MD) simulations, coordinates and momenta of all protein atoms are usually sampled every 0.5-2 femtoseconds over millions even billions time steps. The challenge of MD simulations is to decipher the molecular global motions responsible for the biochemical interests from a huge number of multivariate data. In this seminar, examples will be illustrated through step by step implement of PCA after introducing concepts including covariance, eigenvalues, eigenvector, and principal components. When PCA applied to protein MD trajectories, researchers have found that three or less principal components are normally identified to clarify large scale motions. Applications of PCA in MD simulations of biomolecular systems helps to further understand the microscopic origins of protein motions, which in turn results in a therapeutic benefit to the new drugs design in pharmaceutical design studies.