Comparing Hierarchial, Agglomerative Clustering Methods to Clusters Obtained with Kohenen Self-Organzing Maps

Stefan Chipilov and Carol Parish
Department of Computer Science
Department of Chemistry
University of Richmond, Richmond, VA 23173

Molecular simulations produce a large quantity of data that needs to be mined for physically meaningful results. Conformational searches, molecular dynamics and Monte Carlo simulations on flexible, biologically interesting molecules can produce ensembles containing hundreds or thousands of molecules. There is a great need for methods that can cluster these molecular structures into geometrically related families that provide information about flexibility and dynamical behavior. In this work, we have used three SOM software packages and have also produced several small parsing applications to allow the usage of maestro-formatted data in those same packages. We have performed SOM and XCluster analysis on a series of molecular systems and will present similarities and differences between the results.