Weekly Report(Up until 6th August, 2015)

Anirudh Tiwari

Work Done

Analyzed the interaction strength among varying number of clusters for a given protein by continuously dividing the protein into various clusters by applying k-means algorithm. Steps followed:

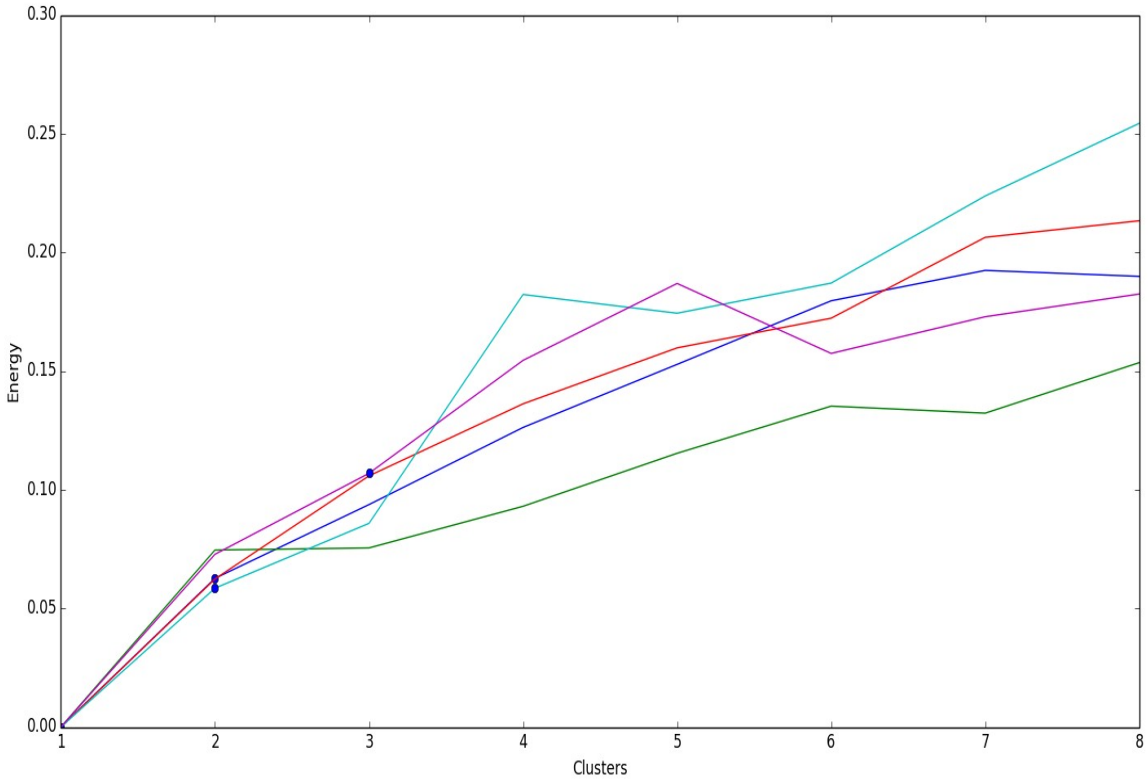
1. For a given protein, apply k-mean by varying k, starting from 1 till 9. Thus, dividing the protein into a single cluster to 9 clusters respectively.

2. For each and every divide, calculate the interaction strength between the various clusters by calculating the inter cluster distance(for those points which are separated by less than 7.0 Å, but belong to different clusters).

3. The total interaction strength is calculated for the entire protein and is normalized by the size of the protein.

Results

It was expected that there should be a steep rise in the overall interaction energy among clusters for a given protein once the number of clusters are more than the number of domains. This was observed in some of the proteins(especially in single domain proteins)



Note: The blue dot represent the actual number of domains for a given protein.

Next Steps

1. Run this script for a bigger set of proteins and analyze the accuracy of the hypothesis.
2. Also analyze the Eigen spectra of the clusters.