REPORT
13.04

Analysis of allosteric communication in TM helical dimers using PARENT

Ruslan Mallaev
LPR, 2022

In this report, I summarize results of a framework I have completed for clustering analysis using data from PARENT.

1 Inclusion by half/all atoms

I checked the difference in the operation of the algorithm in the case when a degree of freedom is assigned to an amino-acid residue only if all corresponding atoms belong to the residue and in the case when at least half of corresponding atoms belong to the residue.

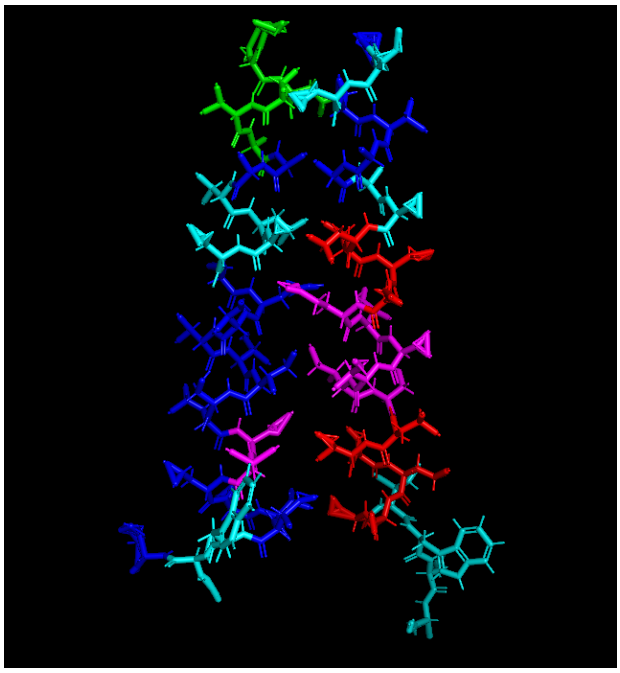

Figure 1: Amino-acid residue clusters identified using "all atom" for v536el2

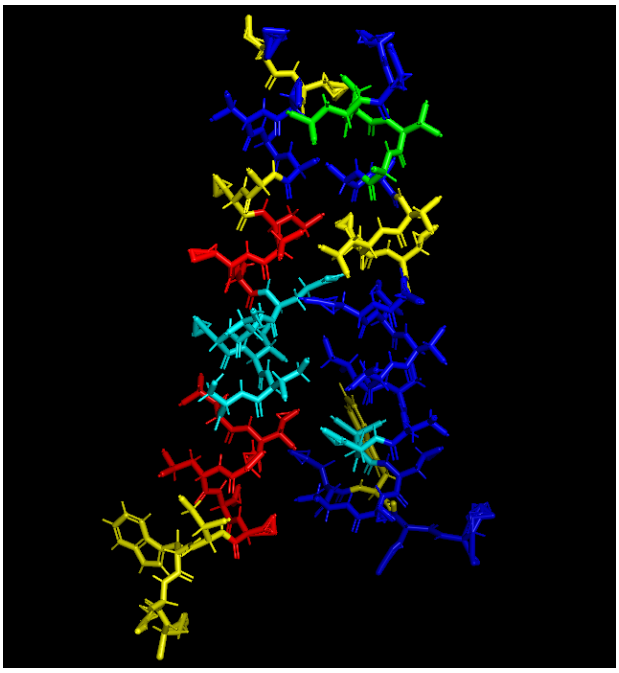

Figure 2: Amino-acid residue clusters identified using "half atom" for v536el2

We decided that it would be better to make the addition using "all atom", since then the "influence" of Backbone is turned off and functional groups play a big role, the interaction of which is most interesting. But the script can also perform calculations for the "half atom" case. For this you need to use the "half/" folder in the appropriate directory when creating "newgroups.txt" file (see previous reports).

2 A metric for estimating the optimal number of clusters

Regarding definition of the optimal number of clusters, it was decided to use two metrics: the sum of external MI normalized by the logarithm of the number of clusters and the ratio of internal MI to external MI. In both cases it was decided to choose the optinmal number according to a change in the first derivative (in first case a local extremum is most often observed at the same point).

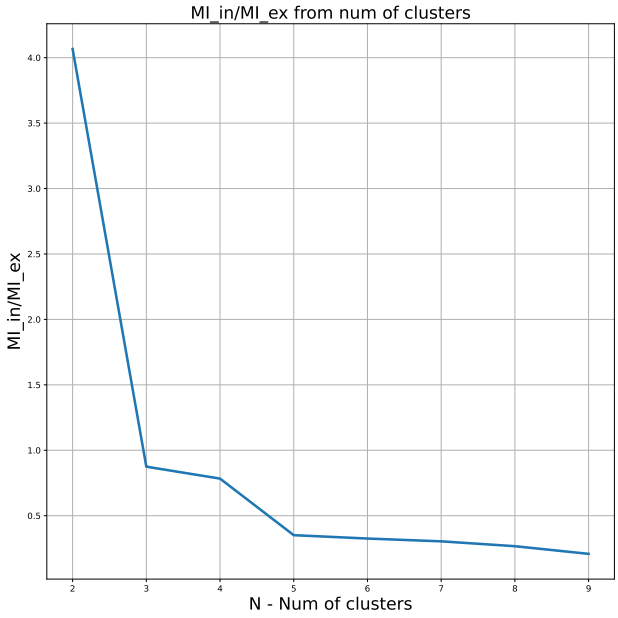

Figure 3: The ratio of internal MI to external MI depending on the number of clusters

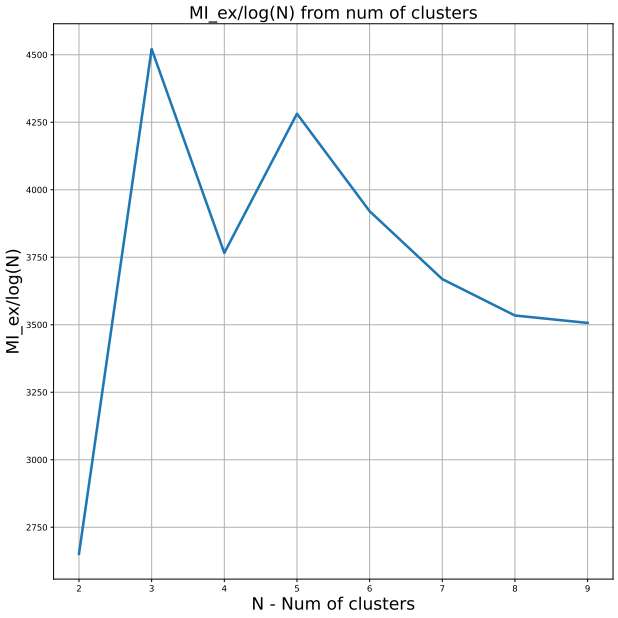

Figure 4: External MI normalized by the logarithm depending on the number of clusters

3 Summary

Thus, the framework is ready to be launched. It was decided to make clustering by amino aced residues (and not by d-d bonds as before), since this has a greater physical meaning. Also, to calculate MI, the residues will include only degrees of freedom, all atoms of which belong to these residues. The metrics presented earlier, on all the structures proposed for analysis, showed themselves quite well. At the moment, the framework is implemented in such a way that the MI similarity matrix for amino acid residues is calculated once and saved to a JSON file, since its calculation takes the most time. Also, the clustering matrices make some sense and now, when using clustering by residues, they look more informative (Figure 5). In such representation color at the intersection of residue numbers means that these two residues belong to a certain cluster colored according to its cluster number as shown with the colorbar.

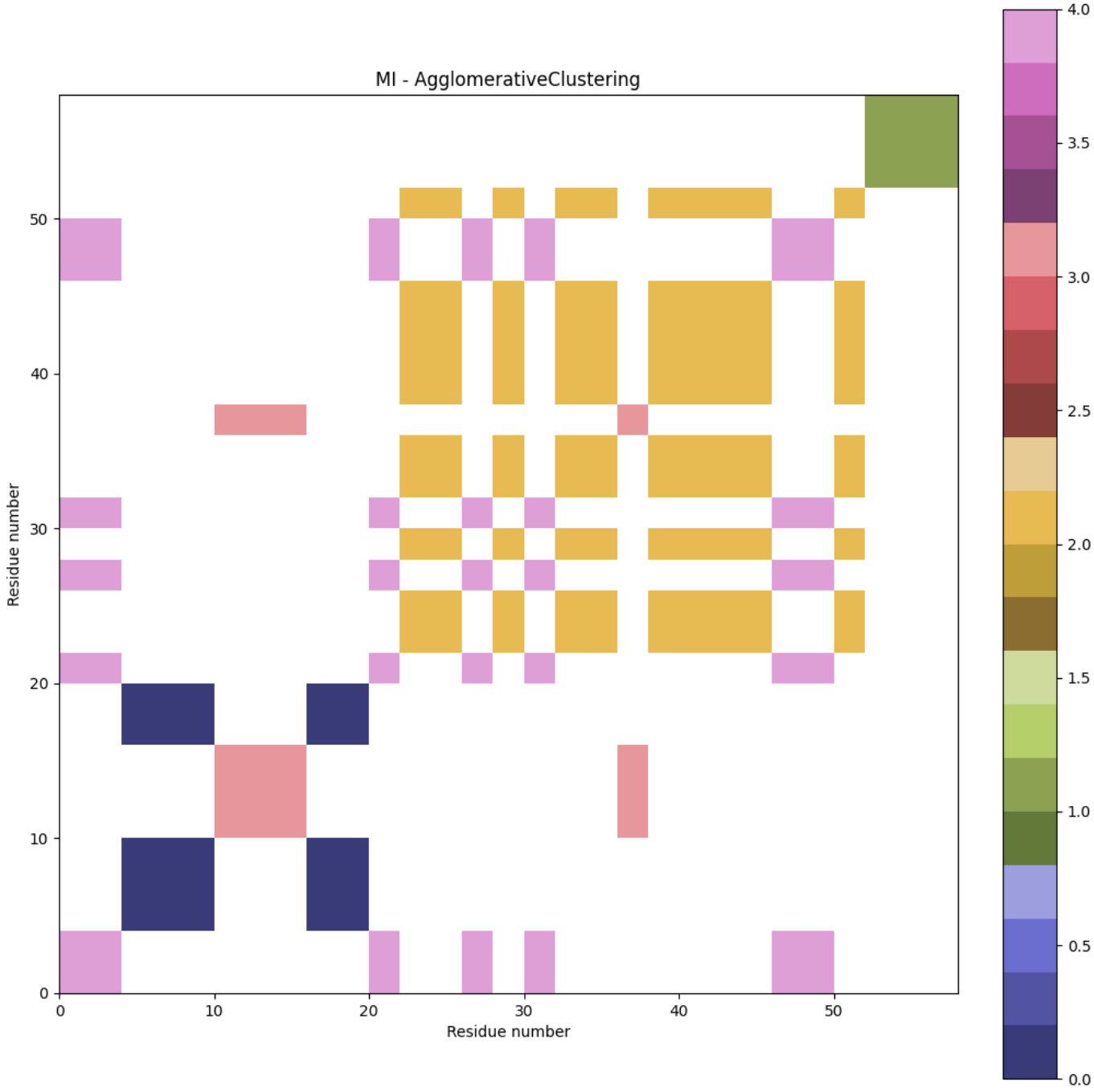

Figure 5: Clustering matrix