

# Visual Analysis of Two Domain Proteins

Anirudh Tiwari

## Work Done

I did a visual analysis of various contiguous and non-contiguous two domain proteins to understand why k-means was yielding inconsistent boundaries of domains and scattered patches of residues.

## Results

I present the case study of 4 proteins highlighting various issues with the k-means output along with possible explanations of the same.

### 1. 1A79(Contiguous)

PDB	Domains	CATH	K-Means
1a79	2	1 A 80 - A 179 - 1 A 9 - A 79 -	83 - 179 & 9 - 82 --



K-means got the boundaries of the second domain as 9-82, while CATH suggest that the boundaries are from residue 9-79. Although residues 80 & 81 are part of the helix which belong to the second domain, hence the Helix is split amongst the two domains which is strange.

2. 1BRY(Contiguous)

PDB	Domains	CATH	K-Means
1bry	2	1 Y 1 - Y 162 - 1 Y 163 - Y 247	22 - 23 25 - 41 67 - 68 160 - 170 182 - 247 & 1 - 18 20 - 21 42 - 66 69 - 159 171 - 181



In this case, it is observed that there are many scattered patches of residues in incorrect domain as per k-means. Upon visualization it is evident that these patches are spatially more close or are closely surrounded by the other domain, hence they are clustered into the wrong domain. Also, there is an error(2 residues) in the boundary, again at a turn.

3. 1AOF(Non-Contiguous)

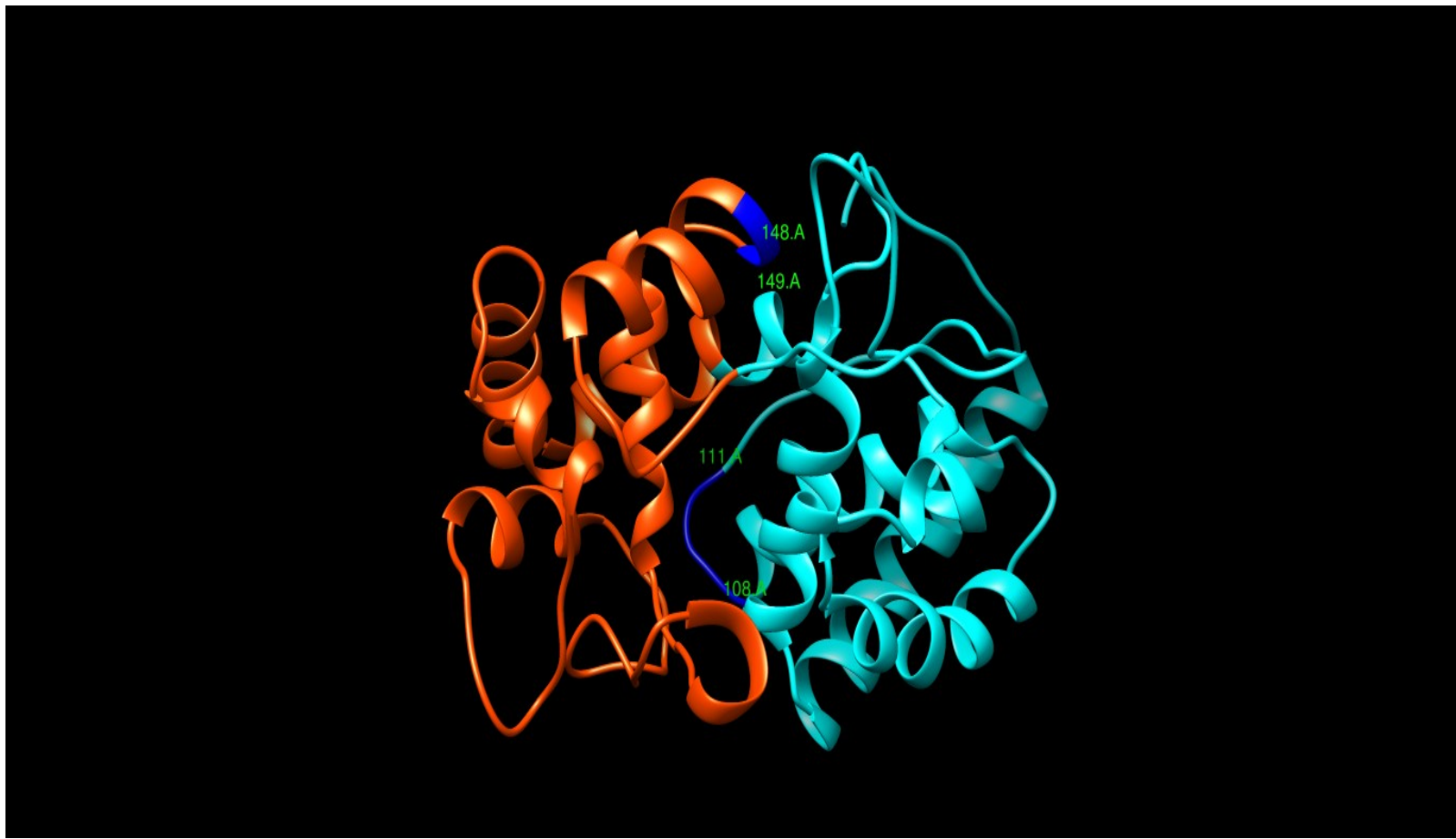
PDB	Domains	CATH	K-Means
1aof	2	2 A 1 - A 82 - A 189 - A 201 - 1 A 83 - A 188 -	1 - 87 189 - 200 & 88 - 152 154 - 188



As per CATH, one of the segment of a non-contiguous domain is from 1-82, while as per k-means it is from 1-87. The residues from 82-87 form a turn which is in between the two domains, rather it forms a bridge between the two domains, thus how to decide how to cluster a turn into one particular domain?

4. 1APX(Non-Contiguous)

PDB	Domains	CATH	K-Means
1apx	2	2 A 2 - A 131 - A 239 - A 250 - 1 A 132 - A 238 -	2 - 33 35 - 107 112 - 130 148 - 149 239 - 250 & 108 - 111 131 - 147 150 - 238 --



Again, small patches are observed which are in vicinity of the other domain, hence are clustered incorrectly. Otherwise the output of k-means is fairly consistent with that of CATH, even though it has non-contiguous domains.

## Conclusions

There are a few points which I concluded post this analysis

1. The domain boundaries need not contain a secondary structure in its entirety, it is possible that a secondary structure(say a Helix) may be divided amongst two domains.
2. It is difficult to correctly judge the boundaries of a domain, as many of the domains have a separating boundary at a turn, with part of the turn on one side and the other part at the other. Thus, it's difficult to draw a line to separate the two domains.
3. K-means showed some scattered fragments because those chunks of the domains were spatially nearer to the other domain or were surrounded by the other domain.