

Domain Decomposition - Ribosome T.Thermo 70S

System:

RNA chains, no protein, one atom per common residue.

335 structures, 3228 atoms (25 atoms were removed prior analysis).*

Similarity (choice):

- distance_fluctuation:
 $M_{ij} = \langle (d_{ij} - \langle d_{ij} \rangle)^2 \rangle$
- distance_range:
 $M_{ij} = \max(d_{ij}) - \min(d_{ij})$

Clustering algorithm:

hierarchical clustering with Ward criterion

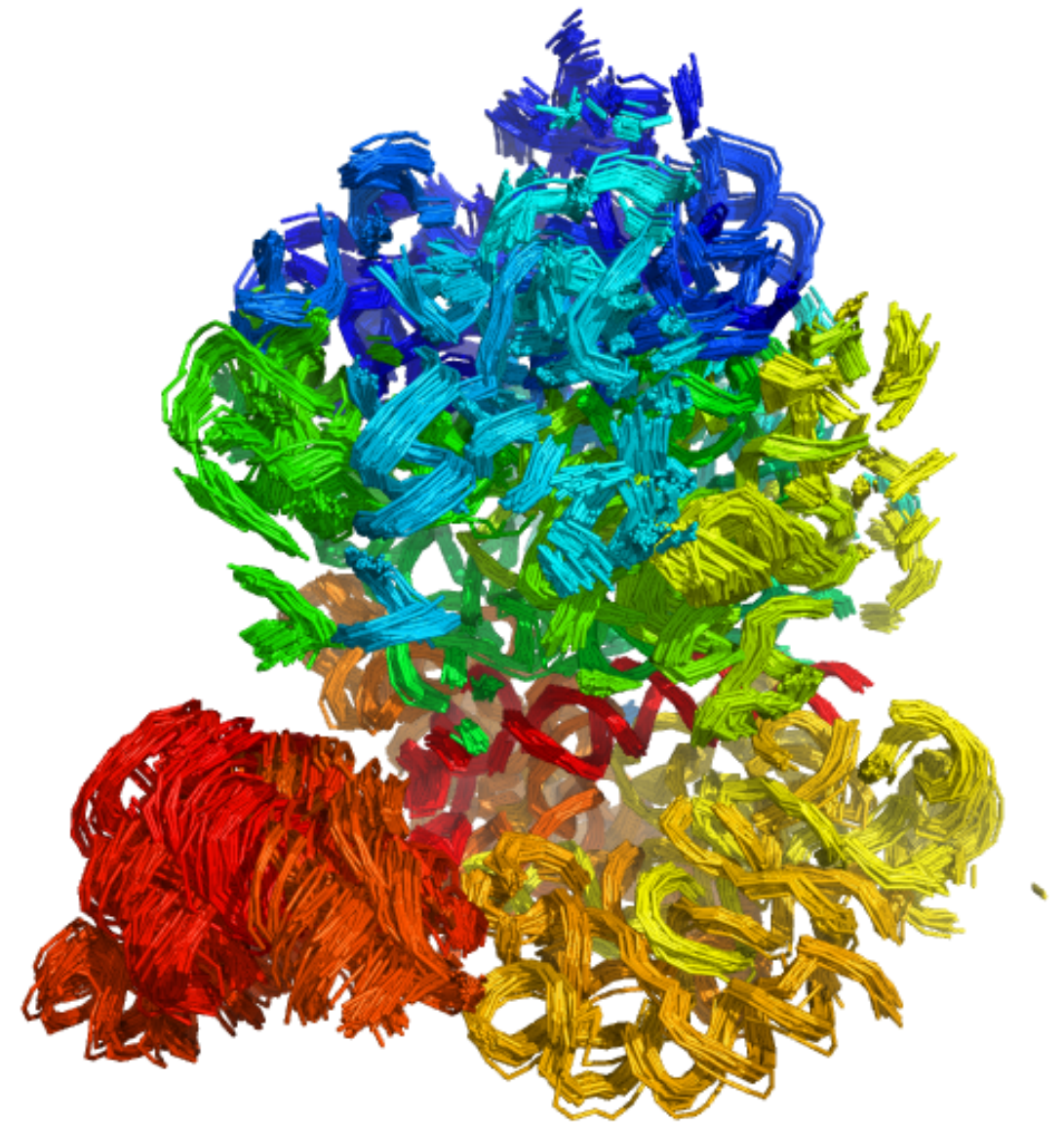
```
pdb_filename = 'data/ribosome_thermo_rna_clean.pdb'
similarity_type = 'distance_fluctuation' #'distance_range' #
frame_range = None #np.arange(0,3,1)
keyword = 'ribosome_thermo_rna/ribosome_thermo_rna_clean'+ '_' + similarity_type

traj = io.load_traj(pdb_filename)
rbclust, rbmat = rb.cluster(traj, similarity_type=similarity_type)
cmpnt.plot_matrix(rbmat, title='similarity', figname=keyword+'_mat')
cmpnt.plot_cluster(rbclust, np.arange(rbclust.shape[0]+1), figname=keyword)

n_clusters = 10
for i_clusters in np.arange(2, n_clusters+1):
    pdb_output = keyword+'_ndom'+str(i_clusters)+'.pdb'
    assignment = rb.get_assignment(rbclust, i_clusters)
    rb.save_cluster_in_bfac(traj, pdb_output, assignment)
```

python script used

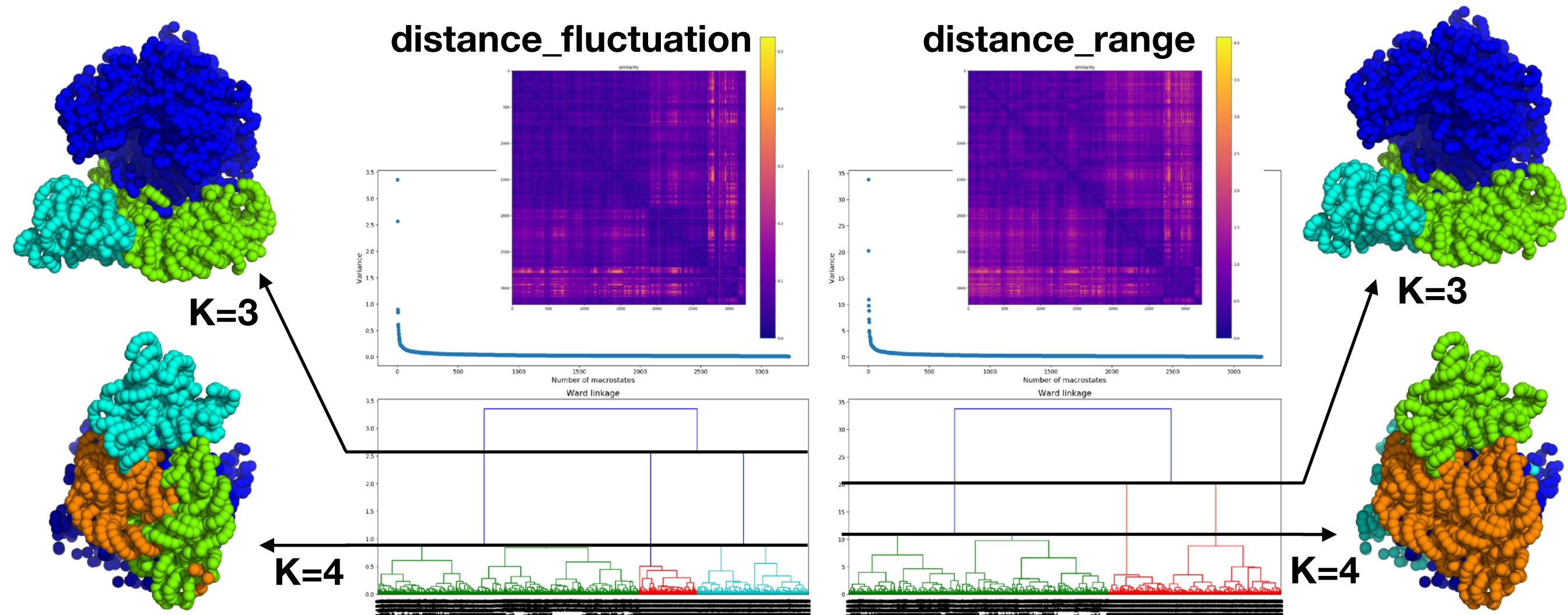
(soon available at <https://github.com/fredericpoitevin/IsCAn>)



[click to download PDB](#)

* the local elastic energy of each atom is computed, allowing to identify outliers

Domain Decomposition - Ribosome T.Thermo 70S



Comments

The similarity matrices look very similar.

Their Ward clustering as well: 3-domain decomposition yields an important drop in total variance, thus suggesting it to be relevant.

An important difference occurs when it comes to 4 domains: using the distance_range similarity matrix decomposes the Large Subunit first, while the distance_fluctuation matrix decomposes the Small Subunit first. The latter decomposition would seem more relevant, as the SSU is where domain closure happens.

Overall, the clustering algo. is a bit messy... (see isolated orange particles above)