### STEP 13 CLUSTERING A SET OF MODELS

The PyRy3D UCSF Chimera Extension lets the user automatically cluster PyRy3D models in order to identify groups of solutions fulfilling restraints. Clustering is essential step of hybrid modeling approach and should be performed after running many independent runs of the program.

Clustering can be also run for models generated with other modeling programs.

We prepared a set of 140 PDB models of 1NIC complex that can be used to test the clustering algorithm.

1. After opening the PyRy3D Chimera Extension window, go directly to the **Clustering** tab.

2. Use the **Input directory** entry field (or the **Browse** button) to select the directory that contains the set of models.

3. Use the **Map file** field to provide the **1NIC** density map file.

4. From the **Score type** menu, choose **CCC.** The structures in each cluster will be listed in ordered based on their fitting to the density map.

5. From the **Measure type** menu, choose **RMSD.** This will be the measure based on which the structures will be joined into separate clusters.

6. From the **Representation** menu, choose **fa**, which means that full atomic structures will be taken into account.

7. In the **Number of structures** entry field, type "140". That's the number of complexes we want to separate into clusters, which is equal to the overall number of complexes we're going to cluster.

8. Set the **Clustering threshold** to 20. This means the threshold in Angstroms that will determine how the models will be divided into clusters.

9. Provide path to the **Output folder**, where the output files will be saved. We recommend you copy / paste the path from **Input directory** entry. so the clustering output files will be saved with the files themselves.

10. Click **Start clustering** to start the process. It might take some time, depending on the number of models and their complexity.
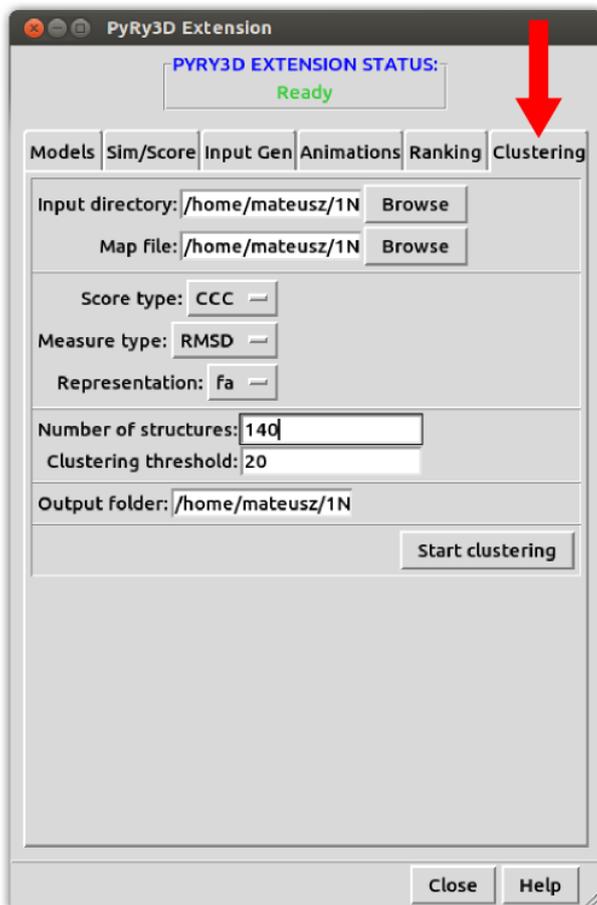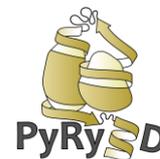
11. The program will generate a set of output files. The most important file is called **clusters.txt.** This file contains names and score values of all clustered models, which are of course divided into separate clusters.