Drug Discovery Project

This report aims to show a practical project I made with Autodock tools, R ggplot2 and basic libraries, and available experimental results on 3CL pro protein from COVID-19.

1. Workflow

The procedure implies docking experiments and experimental and computational data analysis. All of these actions are performed in order to obtain as much information as possible for the posterior discussion about suitable candidates for 3CLpro inhibition.

First of all, some experimental data from enzymatic activity assays is given. In these previous enzymatic experiments, a library of approximately 1000 compounds was tested. The assays served as a screening tool; the highest activities were selected and listed. The selection was done taking account of the deviation in the control diagrams. General workflow for molecular docking and result analysis is represented in figure 1.

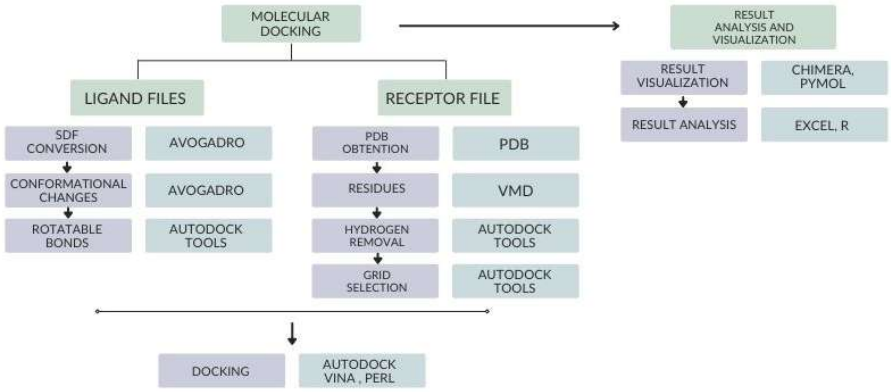

Figure 1. Workflow for the current project that involves molecular docking of a library of 75 compounds on the protein 3CLpro and the following comparison of results with activity that was obtained experimentally. Steps are shown in purple color while tools are in green.

2. Docking results

The first step was the realization of molecular docking simulations with the final 75 ligands listed and its possible conformations and forms. Docking results were obtained with Autodock Vina and a Perl script for multiple ligands docking, poses are shown in figure 2.

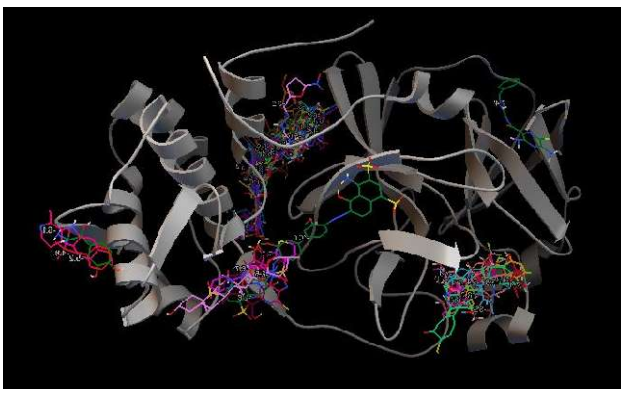

Figure 2. The best docking poses obtained of the 75 compounds in the 3CLpro protein surface.

3. Statistical analysis of results

The statistical analysis of results took account the values of activity percentage for each ligand-protein activity assay and docking affinity. Each possible binding site of the best pose of each compound shown in figure 2 was listed as A, B, C, D or E in one case for different ligand conformations.

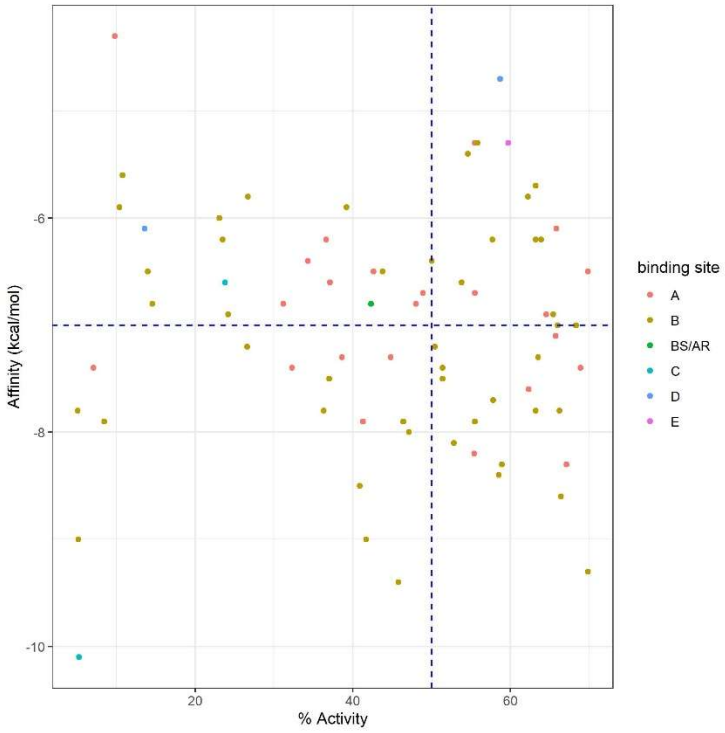

Figure 3. Estimated affinity against activity percentage showing color classification for different binding sites. Thresholds are placed in a mild 50% activity and average activity (-7 kcal/mol).

There's no apparent relation between the binding sites and affinity or activity values. It is shown that binding sites A and B are more numerous, and B is presented more frequently in values with high activity and absolute affinity values. The compound with more than 69% activity and -9.3 kcal/mol activity was a rifamycin derivative.

To relate binding site and activity and affinity values clustering different techniques were used and dendrograms represented. Here are shown the hierarchical agglomerative clusters made with high linkage to show nuances for activity, affinity and activity and affinity values in relation to the binding site position.

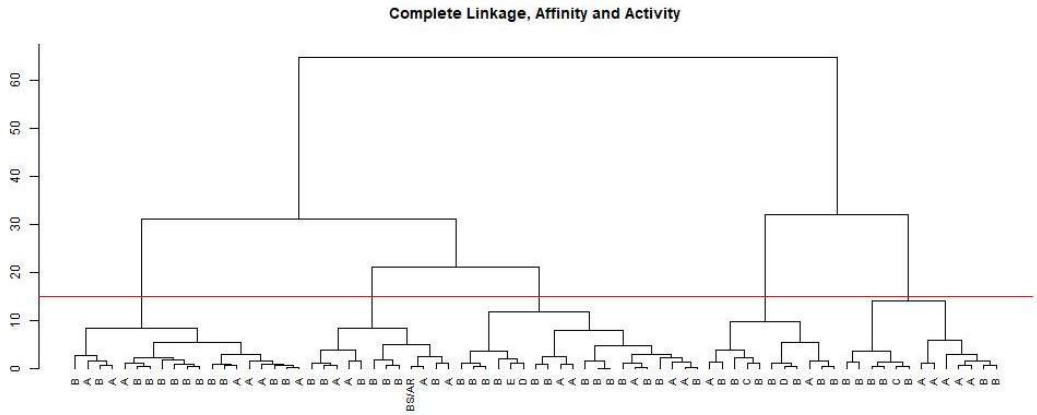

Figure 4. Hierarchical clustering for affinity and activity values and binding site classifications with complete linkage.

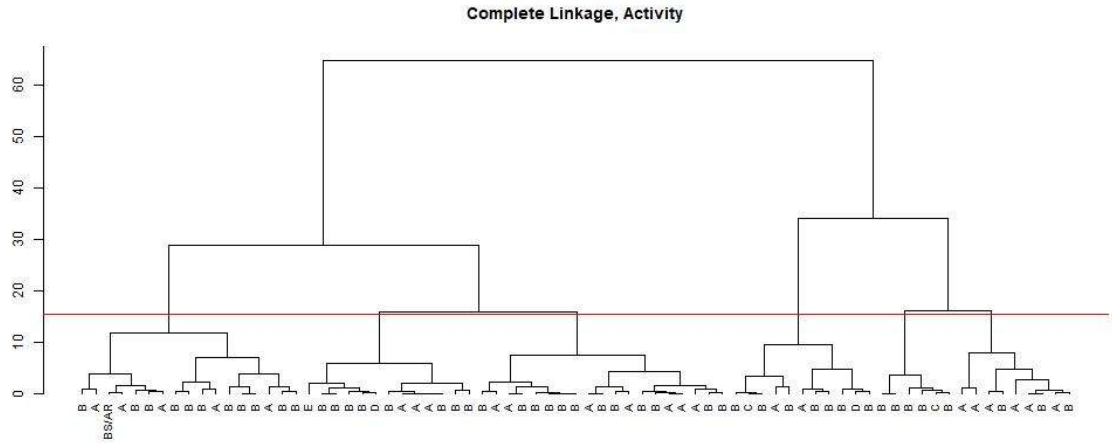

Figure 6. Hierarchical clustering for activity values and binding site classifications with complete linkage.

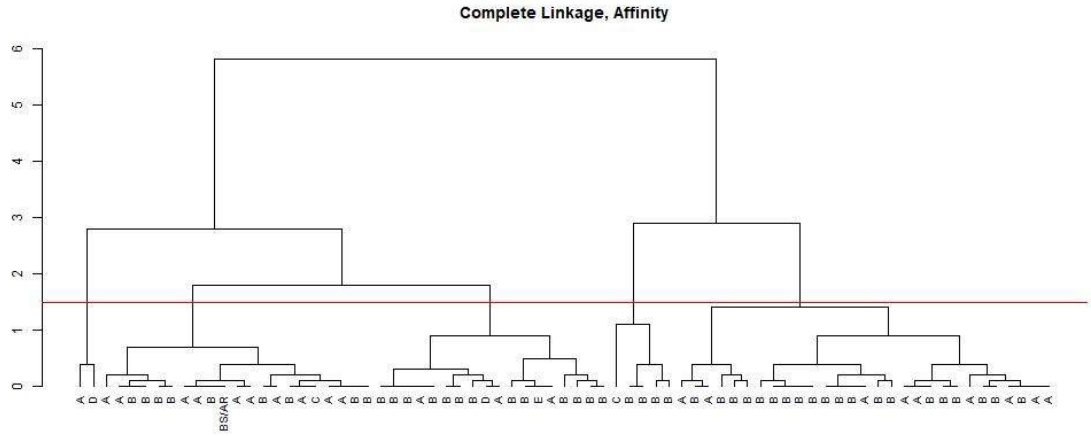

Figure 5. Hierarchical clustering for activity values and binding site classifications with complete linkage.

Definitely no differences are shown among dendrograms for binding site classifications in relation to activity and affinity values given the clusters selected (red line cut in figures 4 to 6).