**SARS-CoV2-CDD-Results**

This project is a computational drug discovery project which attempted to find a compound or a molecule that will inhibit the function of SARS-CoV2 hence its name, SARS-CoV2-CDD, which stands for Severe Acute Respiratory Syndrome Coronavirus 2 Computational Drug Discovery. Chembl Database, a database containing curated bioactivity data of more than 2 million FDA-approved compounds, was the source of the initial bioactivity data for the organism of interest. Replicase Polyprotein 1ab in SARS-CoV-2 was used as the target protein. Replicase Polyprotein 1ab is the key polyprotein that is vital for virus replication as it contains the key proteinases responsible for the cleavages of the polyprotein. Christopher Lipinski's Lipinski descriptors (**Rule-of-Five** or **Lipinski's Rule**) were used to calculate the molecular descriptors for compounds acquired from Chembl. The dataset generated was then split into two bioactivity classes, active and inactive, by a pIC50 threshold value of 6, and tested for statistical significance using the Mann-Whitney U test with a significance level observed at 0.05 P-Value.

Taking a look at pIC50 values, the **actives** and **inactives** displayed ***statistically significant difference***, which is to be expected since threshold values (IC50 < 1,000 nM = Actives while IC50 > 1000 nM = Inactives, corresponding to pIC50 > 6 = Actives and pIC50 < 6 = Inactives) were used to define actives and inactives. Of the 4 Lipinski descriptors (MW, LogP, NumHDonors, and NumHAcceptors), LogP, and NumHAcceptors exhibited ***no difference*** between the **actives** and **inactives**, while the other two descriptors (MW and NumHDonors) shows a ***statistically significant difference*** between **actives** and **inactives**.

It is worth noting that this analysis included several outliers for the features examined herein. Therefore, they may have shifted the statistical power of this experiment, and therefore the differences found may be attributed to them. Since the results of this experiment are very critical, the outliers were not dropped. However, this now prompts for part 3 of this project. For part 3, we will conduct another experiment without the outliers and compare the results obtained from it with those acquired here. In part 4, the outliers will also be studied further to attempt to decipher what they could be and mean.