Meet-U Summary

Structural bioinformatics on the Sars-Cov-2 helicase protein (Nsp-13)

Written by
Amina ALIOUA
Adel BIBI TRIKI
Késia DETHELOT-DELAG
Tiziri TERKMANI

Supervised by
Dr. Vaitea OPUU
Pr. Elodie LAINE
Pr. Juliana SILVA BERNARDES

2023-2024

Summary

The COVID-19 pandemic which is one of the most critical pandemics in human history has highlighted the need for effective therapies. Among the many potential therapeutic approaches, exploring specific viral components shows promise. One target of interest in the fight against COVID-19 is the non-structural protein 13 (nsp13). It is a vital component of the SARS-CoV-2 virus. Understanding and targeting nsp13 could lead to innovative therapies. This approach will contribute to ending the Covid-19 pandemic and preventing future pandemics.

The project Meet-EU therefore fits in this context and could ultimately result in one of the participants finding a new inhibitor for the Nsp-13 helicase. Thus, the goal of our team is to have a final set of ligands that, according to us, could be viable inhibitors.

The key to our strategy was DiffDock, a deep learning diffusion algorithm, because it did both the docking and the scoring, but it also comes with a price. DiffDock is very slow (5 min per molecule in the best case) and using it to dock the whole database would take more than two weeks (non stop). So, our first idea was to reduce the size of the dataset until having a few molecules that we dock. We had a database of 5016 unique molecules (SMILES). Each molecule is transformed into a fingerprint, and more particularly into MACCS fingerprints: Each bit in the fingerprint corresponds to a specific chemical feature or substructure. This makes it easier to understand the molecular characteristics that contribute to similarity. A value of 0 indicates the absence of a character, while 1 shows its presence. We then use Tanimoto distance to compute the distance matrix between all the fingerprints that was used to clusterize the database with the K-means algorithm. Unfortunately, we only had 2 clusters, and after docking a sample from each cluster, we had very close results and we could not have kept only one of them for the next steps. Our teachers suggested that instead of reducing the size of the database, we should try to predict the scores of the molecules that we do not dock. In other words, we randomly pick molecules from the database that we score with DiffDock. Then these scores were used to train a neural network model that predicts the scores for the rest of the molecules in the database. We randomly selected 50 molecules each. We then tested these 100 molecules using Diffdock, compared with the 6ZSL protein, the crystal structure of the SARS-CoV-2 helicase at 1.94 Angstrom resolution.

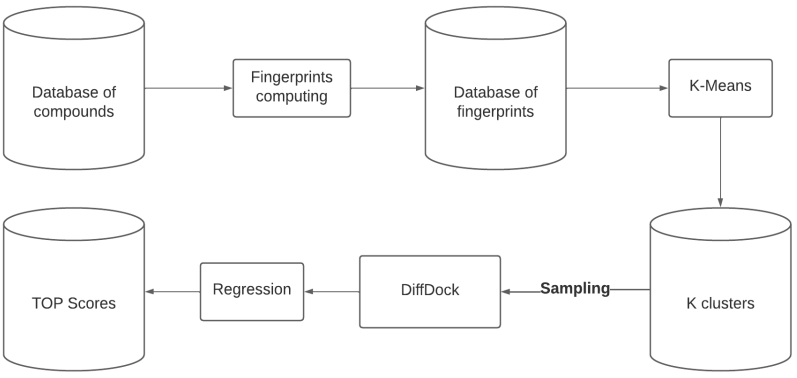
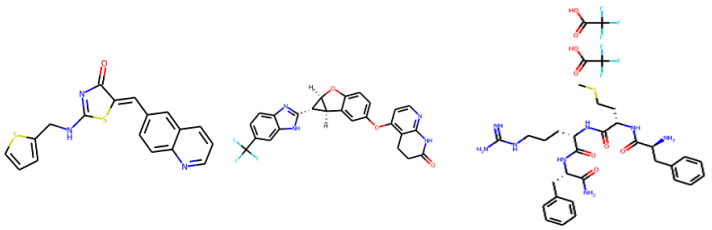


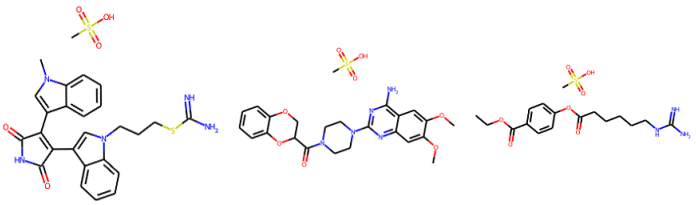
Figure1: Diagram of the full approach

Figure 2: PCA representation of the database after using KMeans

In conclusion, our approach enabled us to predict scores for the 5016 molecules of the database and compare them with some nsp13 inhibitors cited in the literature. As a result, we identified several molecules as potential candidates.



EOS100928 EOS101357 EOS100238



EOS101001 EOS100611 EOS101317

To expand on the topic, we could have followed other tracks. Firstly, we could have used different methods for the clustering, for example other fingerprints, similarity scores and regression models. Thereby, we could have found the most suited ones. Secondly, we could have worked with other features of the database such as hydrophobicity. They could be more relevant regarding our approach. Finally, optimizing the docking by only using the pocket pdb instead of the whole 6ZSL, this aims at reducing computation time and may give better docking scores.