MEET-EU 2-pages report

Stéphane Wang, Alix Bonard, Junwoo Park, Assane Fall, Samuel Lalam

February 9, 2023

Investigating 5014 ligands for potential Sars-CoV-2 helicase inhibitors

Sorbonne University Team3 SAJAS

Methods

A protein-ligand interaction can be tested in 3 steps. First the preparation of ligands, second the preparation of the protein and third execution of the pipeline(Fig.1).

Ligand preparation

To process ligand screening we had to generate the ligand 3D structure. For this aim, we used Rdkit in python. 5014 ligands from the European Chemical Biology Library(ECBD) were screened as potential inhibitors of the nsp13. The 2D and 3D structures can be obtained with the Simplified Molecular Input Line Entry Specification (Smiles) metric. And we added hydrogens for each generated ligands to further uses of ligand in docking steps. It is also possible to generate multiple conformations of ligands in the Rdkit, however it is an unnecessary step since Autodock-vina uses its own conformation change algorithm to decide the lowest free energy of inter-molecular levels. And at last, ligand stabilization is done with the Merck Molecular Force Field(MMFF94) that is a family of chemical force fields developed by Merck Research Laboratories.

Protein preparation

We selected the 3 protein crystallographic structure of nsp13 which are 6ZSL, 7NIO and 7NN0. In these protein files, there are also ligands and H20 which are already in contact with the protein, so we need to eliminate these non-substantial residues and after, we proceed the free energy minimization step using Amber force field in Chimera. For the standard residues we used AMBER, ff14SB and AM1-BCC for the other residues. The crystal structure of 6ZSL is in 1.94 Angström resolution.

Pipeline

We developed an automatic pipeline that generates a report.csv file containing the interaction score of ligands with detected pockets. First, once a protein and ligands are prepared, the p2rank(ML based method) detect all the possible pockets in the given protein, and then, the ligands are automatically docked using Autodock-vina. And lastly, convex-pl scores all pocket-ligand interaction and these scores are recorded in the report.csv file. The average execution time for a ligand-pocket pair is about 1min in the Intel i5 cpu without gpu system. All the parameters of the three tools can be modified in the params.txt before the starting of pipeline. In this simulation, the size of the docking box which is here x = 20, y = 20, z = 20. The time to test a ligand in our pipeline is around 1 minute. Yet, for the 5014 ligands tested, we took 4 days to obtain all scores for the pocket with the highest existence probability.

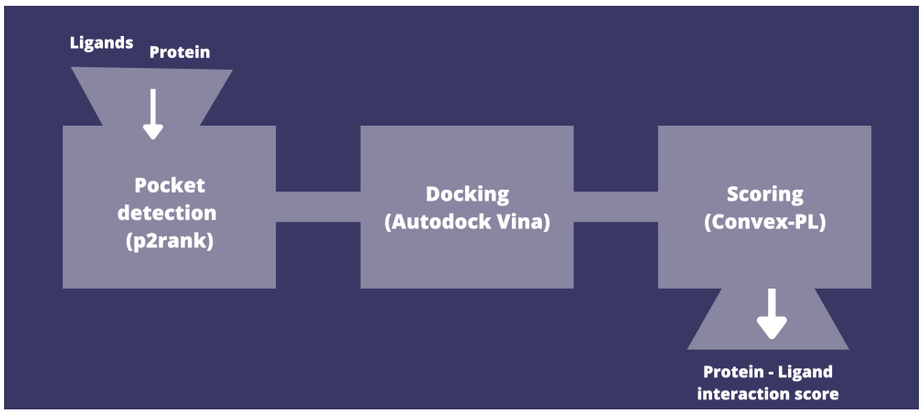
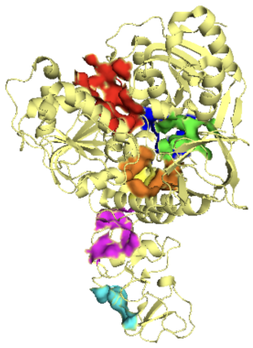


Figure 1: Pipeline flowchart.

Results

Pocket detection

First, we explored pockets on the SARS-CoV-2 helicase structure (nsp 13 PDB:6ZSL) using machine learning methods to detect probable pockets. We obtain seven possible pockets with variable probabilities. The highest value is 0.786 and corresponds to the pocket 1(Fig.2). We also observed a drop in the probability, passing to 0.786 to 0.261 and going less and less for the pocket 3, 4, 6 and 7. Considering those results, we decided to choose the pocket 1 to screen the 5014 ligands for our first approach. This pocket has already been discovered as an ATPase pocket.



name	rank	probability
pocket1	1	0.786
pocket2	2	0.261
pocket3	3	0.100
pocket4	4	0.069
pocket5	5	0.055
pocket6	6	0.032
pocket7	7	0.031

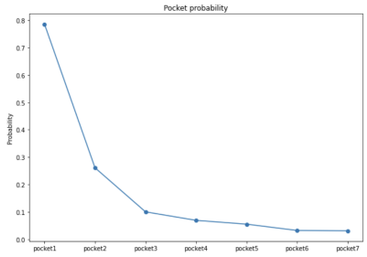


Figure 2: Overview of possible seven binding pockets detected on the nsp13 protein (PDB:6ZSL) using machine learning method p2rank. Identification of seven pockets(left), pocket 1 (red), pocket 2 (green), pocket 3 (blue), pocket 4 (yellow), pocket 5 (magenta), pocket 6 (cyan), pocket 7 (orange). Table presents the probability of existence of a pocket and the rank of the pocket detected depending on the probability value.

Score and comparison

Top 10 ligands - Best ligands - Conformations

We present 10 ligands with the highest score over the 5014 ligands from the database. We compared those scores to the score of three ATP conformations as we examined the ATPase pocket. We only examined a score higher than the first ATP conformation (Fig.3, ATP_conf0_0) as we search for ligands that can bind the ATPase pocket as well as the ATP. Moreover, in terms of probability, taking the highest score into account could enable us to have the most probable ligand suitable for the pocket. The highest score obtained is 28.902 for the EOS100851 ligand in a certain conformation. This ligand is also known as Grazoprevir. This drug component has been already identified as a inhibitor of the SARS-CoV-2 helicase in vitro and is a FDA-approved drug.

Nsp13 - grazoprevir interaction

From the Fig.4, we can check the hydrophobicity and the electrostatstic surface of the nsp13 and it also shows the possible hydrogen bonds between nsp13 pocket1 and the grazo-

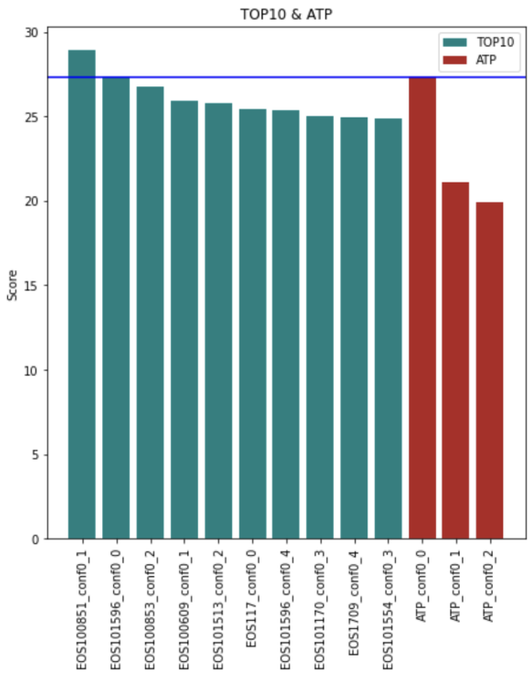


Figure 3: Top 10 ligands from the 5014 ligands using pipeline, blue threshold is the score of ATP on the pocket.

previr. There are 6 possible hydrogen bonds and 5 for the oxygens of the sulfur part.

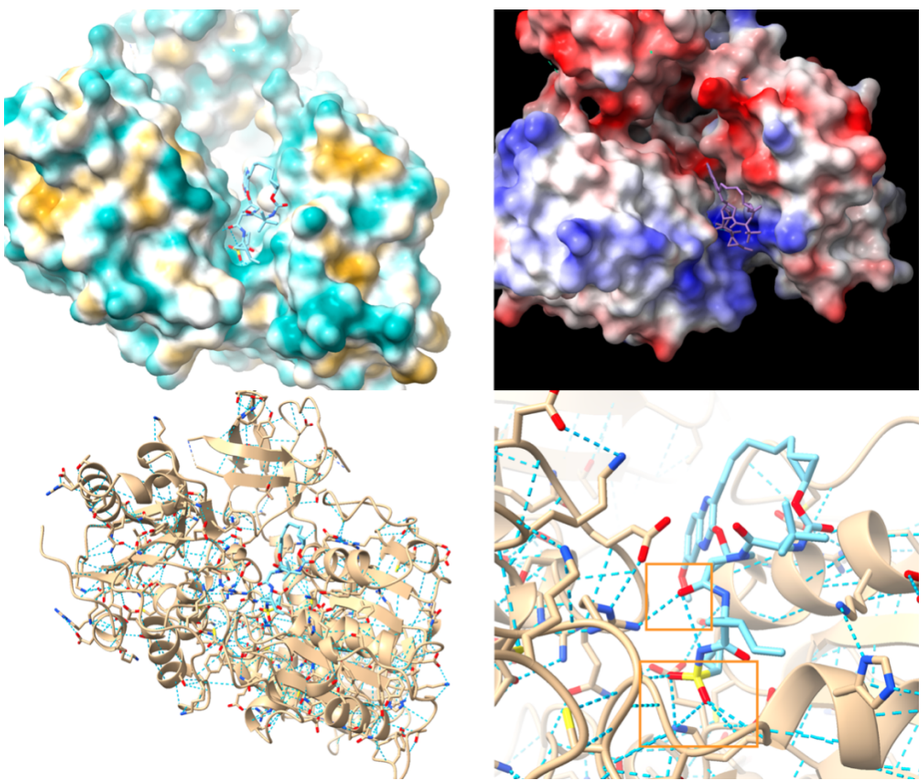


Figure 4: Hydrophobicity(left top) of the pocket with the docked ligand(grazoprevir) and the electrostatic surface(right top) of a protein. Global view(left bottom) of nsp13 and the ligand(grazoprevir) and the close up view(right bottom) with hydrogen bonds(oragne rectangular), 6 possible hydrogen bonds between nsp13 and the grazoprevir.