

MEET-EU project

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

Abstract

The SARS-CoV-2 virus causes an infectious disease which can induce mild to moderate respiratory illness. Yet, some infected people could experience very serious illness with the requirement of medical attention. In 2020, the SARS-CoV-2 has spread to all continents creating a pandemic. Nowadays, 7.55 millions cases of infection have been confirmed[7] which constitutes a real sanitary issue. Thus, therapeutic targets are needed to slow infections, and in the long term, stop the worldwide propagation of SARS-CoV-2.

Introduction

The SARS-CoV-2 is a single-stranded RNA virus. The interaction of the Spike protein and the cell surface receptor ACE2 mediates the membrane fusion. From the viral genome, the cell ribosomal machinery produces 2 proteins pp1ab and pp1b. Cleaved, those proteins generate approximately 16 non-structural proteins (nsps) and 10 structural proteins. Structural proteins consist of the envelope, membrane, nucleocapsid and the spike proteins and the non-structural proteins carry the viral replication activity. Replication and transcription complex (RTC) of the SARS-CoV-2 is composed of the nsp7, nsp8, nsp9, nsp12, nsp13 and nsp14[10]. The role of some nsps remains unclear. However, it has been shown that the RTC activity could be supported by the association of one nsp7, two nsp8, one nsp12 and two nsp13[19]. Indeed, the nsp12 is identifiable as a RNA-dependent RNA polymerase (RdRp) combining with nsp7 and nsps8, and the nsp13, as an helicase.

One of the main issues of this pandemic was to find a vaccine to stop the viral infection no matter the variant. For this purpose, several drugs were studied and tested to target SARS-CoV-2 proteins. An example of protein candidates are the Spike or the RdRp protein[8]. Nonetheless, other proteins can be considered regarding the life cycle of the virus. One interesting potential candidate is the nsp13. It is a highly conserved protein among the coronaviridae family, which makes the nsp13 a robust candidate for a vaccine suitable for variants regarding its low mutation rates within coronaviruses[12]. Moreover, this protein has pleiotropic functions, one of which is the helicase activity. Characterized as part of helicase superfamily 1B, the nsp13 varies from the SARS-CoV-1 by one residue[18]. Crystallography studies revealed that the nsp13 structure is composed of five domains. Two ATPase domains, RecA1 and RecA2, a N-terminal zinc-binding domain (ZBD) enabling the interaction with RTC and a 1B domain. From ATP hydrolysis, this helicase allows the unwinding of single-stranded nucleic acid molecules with a 5'-3' direction. Associated with the RTC, prior studies suggested that nsp13 could facilitate and regulate the RdRp activity[4]. Thus, the nsp13 inhibition could limit virus infection.

Prior studies identified seven nsp13 pockets using molecular dynamics simulation(MD), of which an ATP pocket, a RNA pocket, a ZBD pocket, a Stalk pocket and three RecA2 pockets[3].Regarding the nucleic acid unwinding and the NTPase activity, the ATP and RNA pockets were the most targeted and analyzed in vitro as well as in silico. For example, it has been found that licoflavone C[6] inhibits both nucleic acid unwinding and the NTPase activity as well as SSYA10-001[17] which inhibit in a noncompetitive way SARS-CoV helicase. In this report, we investigate potential nsp13 inhibitors by exploring nsp13 pockets and by screening

5014 small molecules from the European Chemical Biology Library. For this purpose, we develop an automatic pipeline enabling pocket detection, docking and the scoring interaction between a potential ligand and a target nsp13 pocket. We considered different approaches to inhibit the nsp13 by (1) finding the best ligand for the highest probability pocket, (2) finding a ligand that can interfere with the nsp13 - nsp8 interaction.

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.3).

Ligand preparation

To process ligand screening we had to generate the ligand 3D structure. For this aim, we used Rdkit[16] 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[16], 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)[11] that is a family of chemical force fields developed by Merck Research Laboratories.

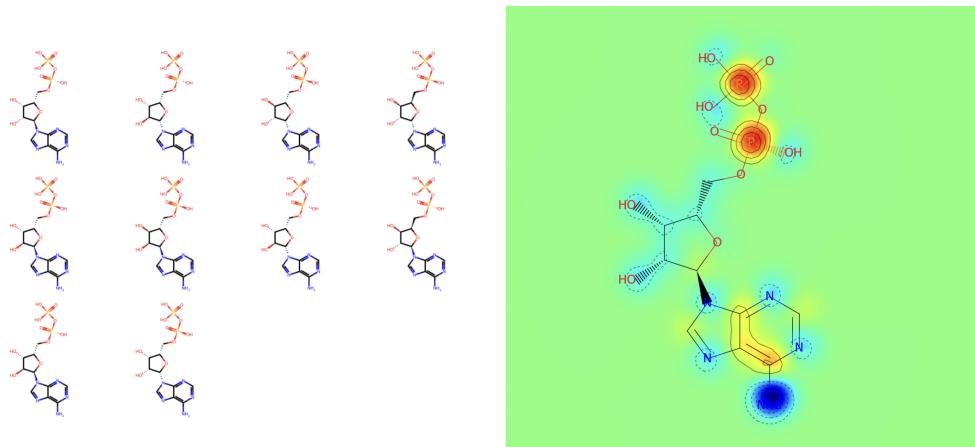


Figure 1: 10 generated conformers of a ligand using Rdkit in python(left) and its partial charges(right)

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 H2O 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[1], ff14SB and AM1-BCC for the other residues. The crystal structure of 6ZSL is in 1.94 Ångström 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)[15] detect all the possible pockets in the given protein, and then, the ligands are automatically docked using Autodock-vina[2]. And lastly, convex-pl[5] 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

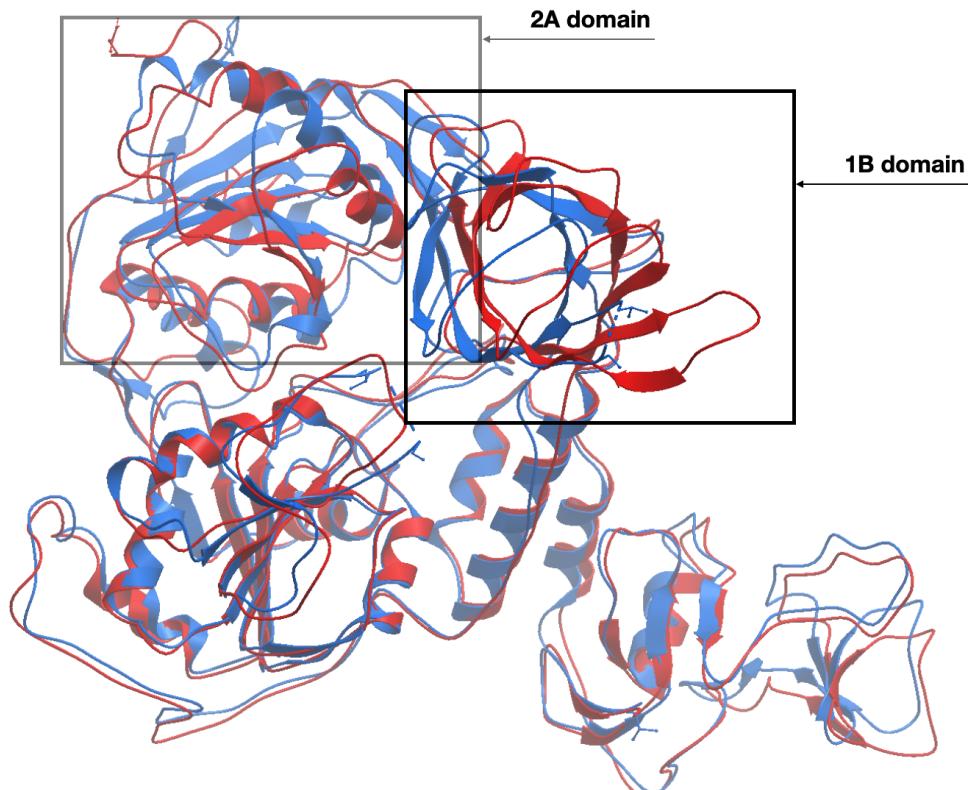


Figure 2: Nsp13 in APO structure before(blue) and after(red) the minimization of free energy.

the 5014 ligands tested, we took 4 days to obtain all scores for the pocket with the highest existence probability.

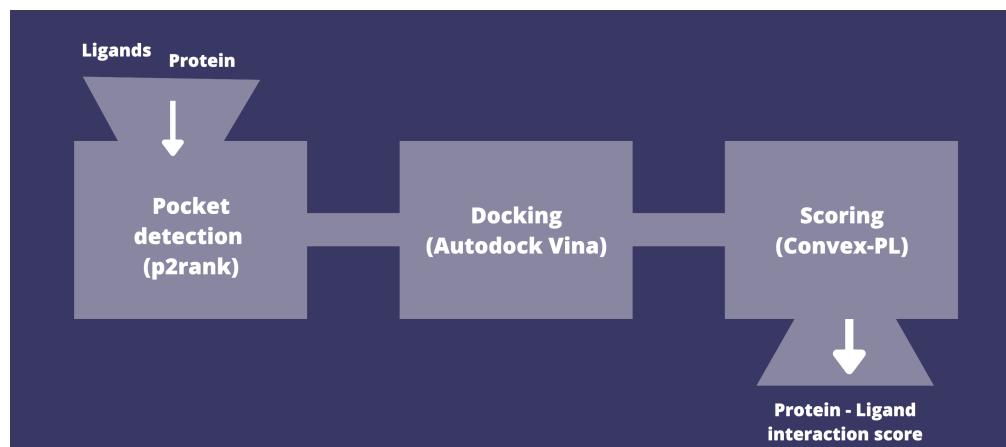


Figure 3: Pipeline flowchart.

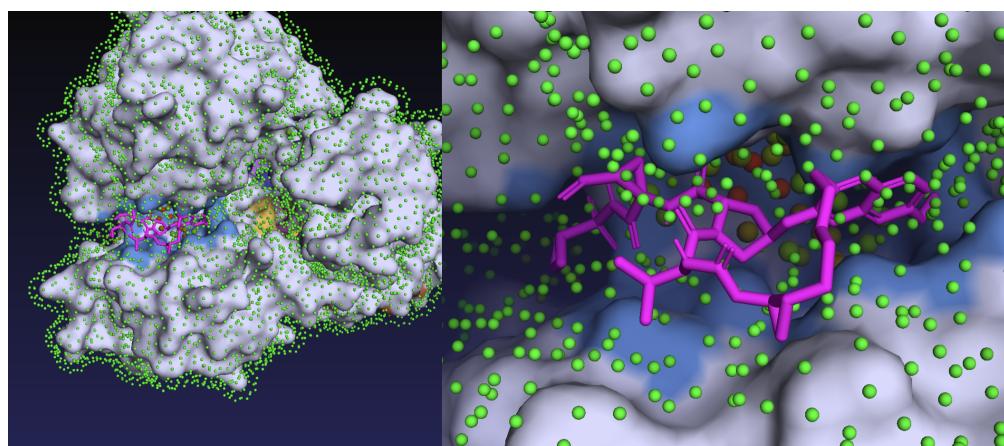


Figure 4: Detected pocket in nsp13 with a docked ligand. The dots around the surface indicate the probability of pocket and all possible pockets are colored on the surface. Probability of pockets are from 0(green dot) to 1(red dot)

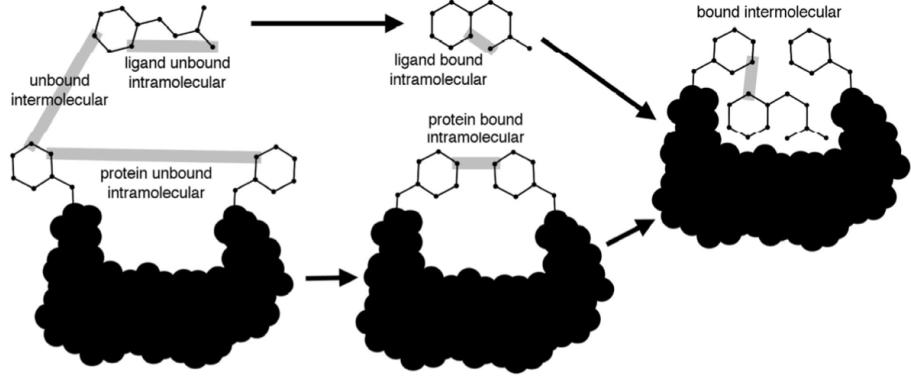


Figure 5: Simplified flowchart of Autodock-vina[2] docking process.

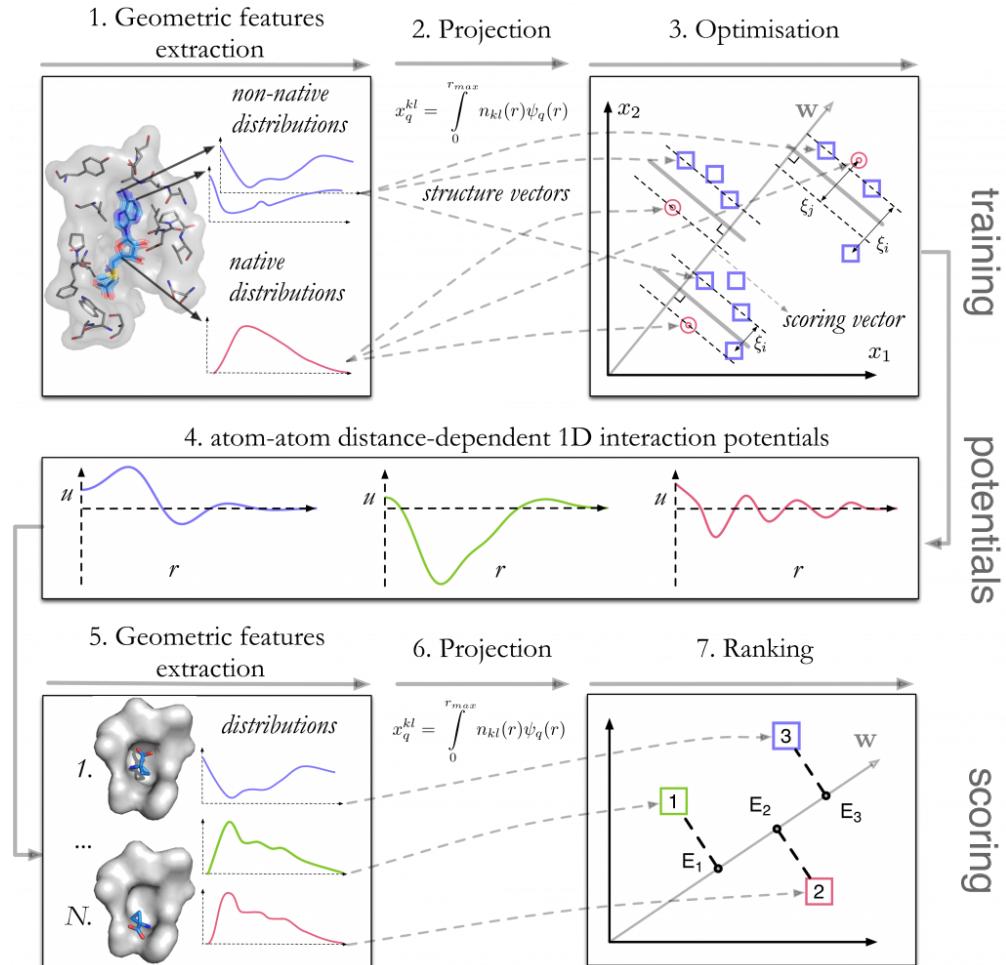


Figure 6: Convex-PL[5] flowchart.

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.4). 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[4].

For the second one, the pocket corresponding to the interface between nsp13 and nsp8[19] has a probability of 0.032. Visualization of the interaction nsp13 - nsp8 takes place at the surface where no pockets are identified visually for all nsp13 conformations (Fig.7). Despite the very low probability of finding a pocket, we also considered this pocket. Indeed, for this part we considered the nsp13 in the RTC context. The SARS-CoV-2 helicase is critical for the RTC activity. It has been found that the nsp13 bound the RTC after the binding of

nsp8 and 7 on nsp12[19]. Blocking the nsp13 - nsp8 interaction could prevent the functional RTC complex.

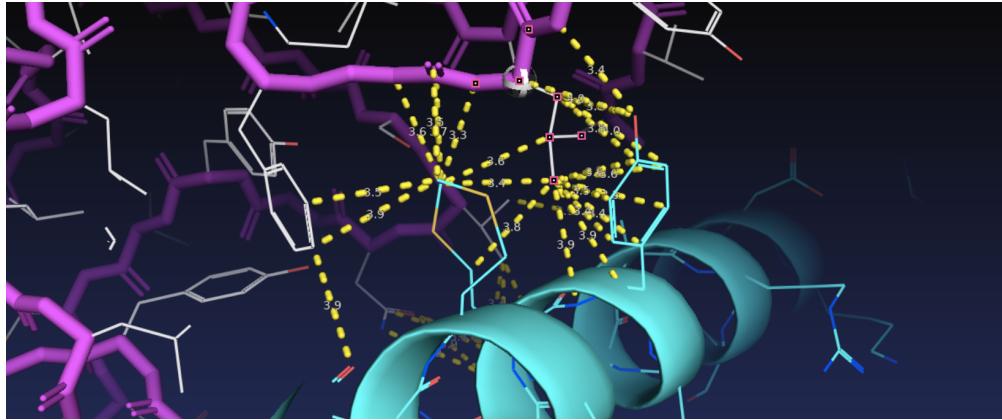


Figure 7: Blue:nsp13, magenta:nsp8.

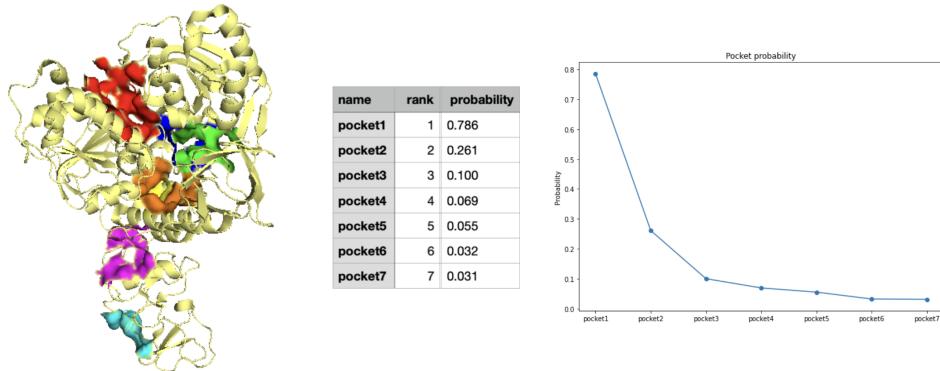


Figure 8: Overview of possible seven binding pockets detected on the nsp13 protein (PDB:6ZSL) using machine learning method p2rank. Identification of seven pockets(left), pocket 1 (red), pocket 2 (green), pocket 3 (blue), pocket 4 (yellow), pocket 5 (magenta), pocket 6 (cyan), pocket 7 (orange). Table 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.9, 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[8].

To go further, we tested 30 different grazoprevir conformations. We observed a maximum score of 39.542 (Figure 3. A). It could be interesting in further studies to analyze whether this conformation could be naturally stable in the ATPase pocket.

In the Fig.11, we can check that the suggested ligands from Heidelberg universities show the good scores on the pocket1.

One of the aims of the project was to compare our results with teams from the University of Milan and from the University of Heidelberg. In Fig.11, we present our top 10 ligands and the top 10 ligands of each team. The score of each ligand was determined through our pipeline.

The team from the University of Heidelberg selected the drug-like molecules from the Zinc database[9]. Pocket detection was performed using Fpocket. Top 10 ligands were docked using AutoDock Vina and scored using Virnardo scoring function. They validate their results using Molecular dynamics using GROMACS. They also did ligand optimization using

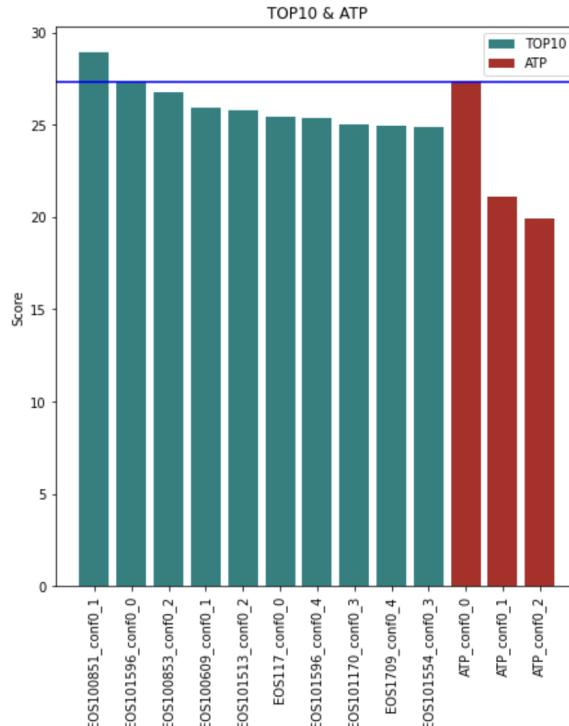


Figure 9: Top 10 ligands from the 5014 ligands using pipeline, blue threshold is the score of ATP on the pocket.

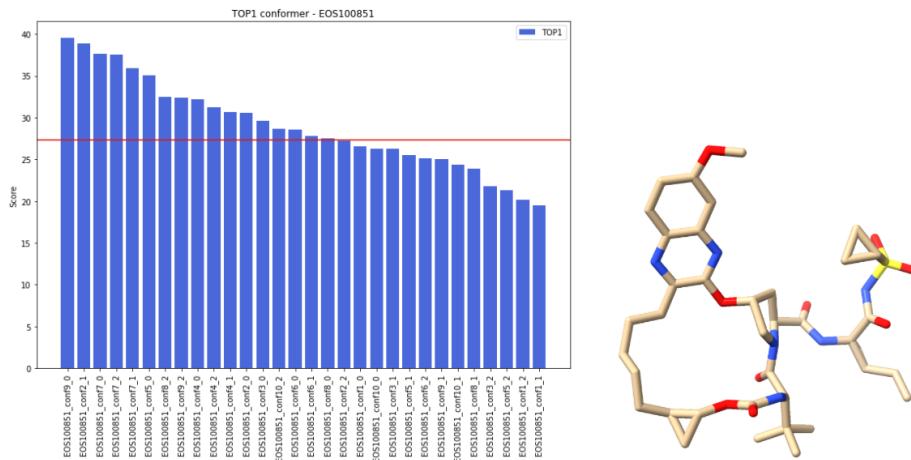


Figure 10: TOP 30 Grazoprevir conformations score using Convex-PL. Score of 30 grazoprevir conformers EOS100851(left). A threshold here corresponds to the ATP score of 27.309). 3D Visualization(right) of the best grazoprevir conformer with oxygens(red), nitrogen(blue), sulfur(yellow) and carbon(beige)

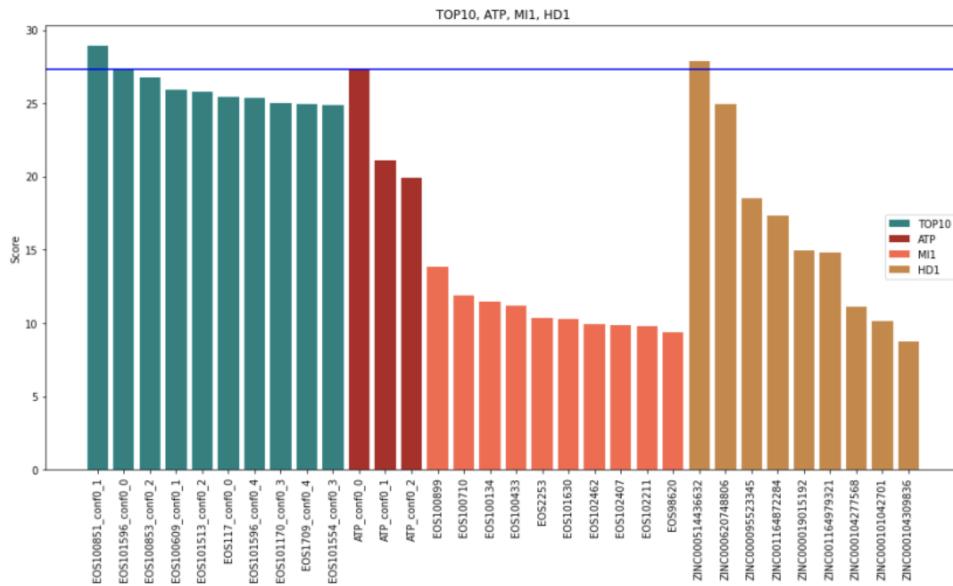


Figure 11: Comparison of the TOP10 ligands on the pocket1(Fig.8) with the universities of Heidelberg and Milan.

AutoGrow4 before docking simulation. The crystallographic structure of the nsp13 is the same as we used in our project (PDB:6ZSL) and the same pocket (ATPase pocket). The team from University of Milan used a different nsp13 crystallographic structure (PDB:5RM2). This team cleaned and standardized ligands from the ECBD database by removing molecules not drug-like for example, before screening them. This team performed machine learning methods to determine the top 10 ligands. They performed a docking simulation of their training set composed by active and inactive known inhibitors of the nsp13 using the software PLANTS. Scoring results were obtained with Chemplp and Rescore+. This team considered two pockets (ATPase and the target pocket $x = -12.95$, $y = 39.96$, $z = -18.27$) to create their model equation. Once this model defined, they screen the cleaned and standardized ligands. We observe that the top10 ligands of the University of Milan have low scores compared to our results and University of Heidelberg, one. The Milan ligand with the highest score corresponds to the phenytoin molecule, an anticonvulsant. For the Heidelberg ligand with the highest score, we obtain the ZINC000514436632 molecule. Compared to our result and the ATP score, we suggest that the top10 ligands from the University of Milan are not specific to the ATPase pocket, regarding the construction of their model based on two pockets. The University of Heidelberg has a better score than the ATP score, so it can interact with the ATPase pocket residues. However, compared to the grazoprevir, few studies considered ZINC000514436632 as a possible inhibitor molecule.

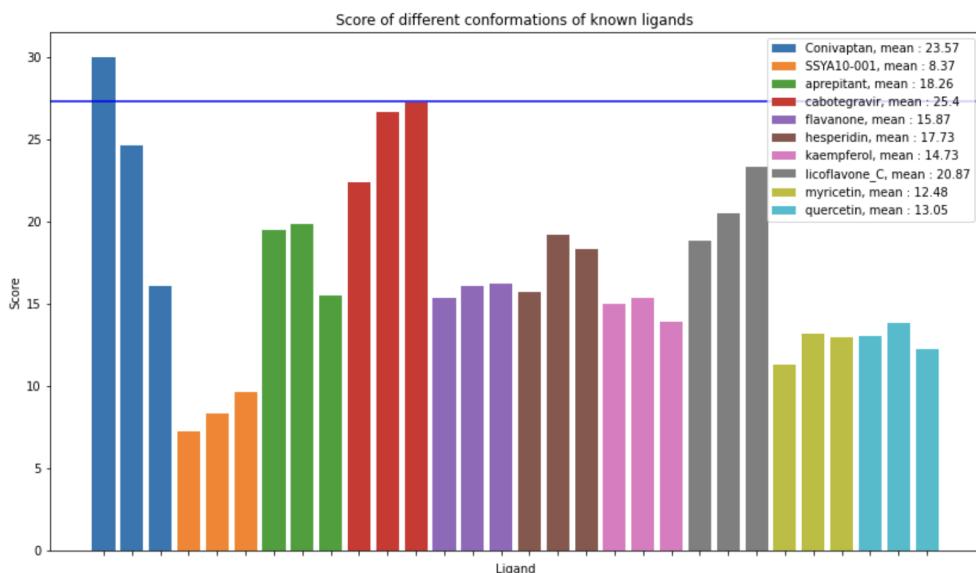


Figure 12: Scores of known ligands to pocket 1 (Fig.8), blue threshold is the score of ATP on the pocket.

Next, we generated scoring results¹² from known ligands to know how our pipeline fared and scored them. This step is important to interpret the specificity of the ligands that we found. Indeed, SSYA10-001 has a low score with our pipeline and that can suggest a poor interaction with this pocket. But, it has been proved that this molecule is a noncompetitive inhibitor and an effective SARS-CoV inhibitor. Regarding these results, we can thus propose that high-scoring ligands in our pipeline might not be noncompetitive ligands.

Nsp13 - Nsp8

We tried to find a ligand that could inhibit the RTC activity by blocking the nsp13-nsp8 interaction. We noticed from the results shown on Fig.1S that some of the residues involved in the interaction between the two nsps also made up the sixth pocket obtained from the pocket detection step. Despite the low probability of the pocket (0.032), we still tried to look into the interaction of possible ligands with the pocket. We generated 5 conformers for each 100 randomly selected ligands from the database and ran them through the pipeline to obtain their interaction score. We show on Fig.13 the ligands with the highest interaction score with this pocket. However, since the pocket is of low probability and we lack data, we are not able to compare the score obtained. Thus, we can only assert that the pipeline is able to dock and calculate an interaction score for this pocket. It would be interesting if we had an interaction score between nsp13 and nsp8. Moreover, we were not able to test all the ligands from the database since the ligands were tested on all the detected pockets.

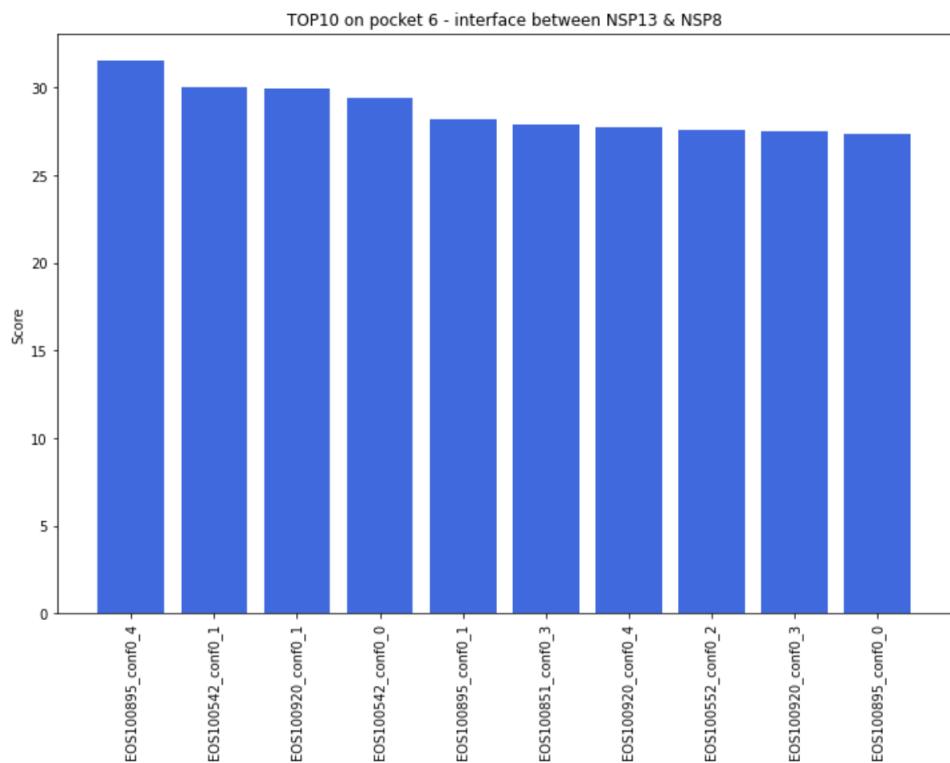


Figure 13: Ligand scores on the pocket6.

Nsp13 - grazoprevir interaction

From the Fig.14, we can check the hydrophobicity and the electrostatic surface of the nsp13 and the Fig.15 shows the possible hydrogen bonds between nsp13 pocket1 and the grazoprevir. There are 6 possible hydrogen bonds and 5 for the oxygens of the sulfur part.

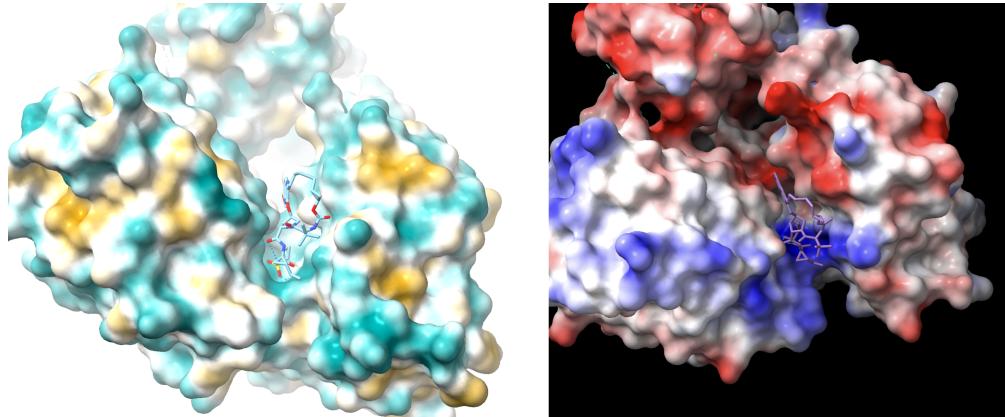


Figure 14: Hydrophobicity(left) of the pocket with the docked ligand(grazoprevir) and the electrostatic surface(right) of a protein.

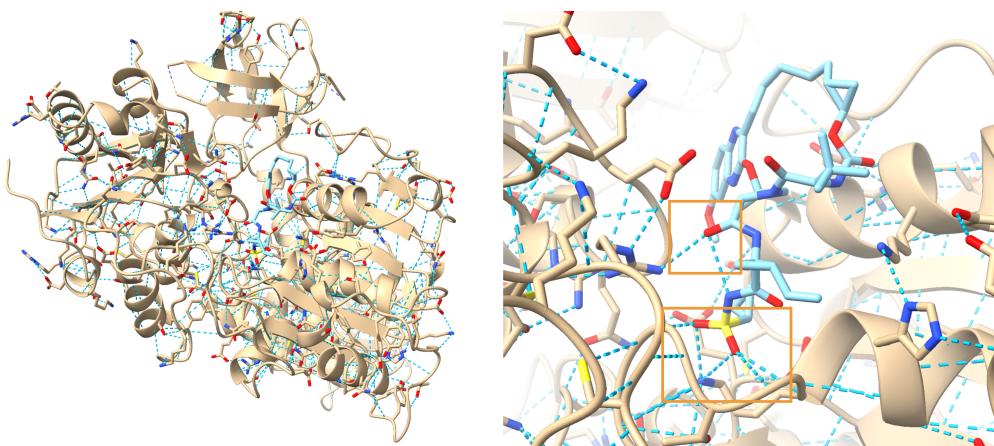


Figure 15: Global view(left) of nsp13 and the ligand(grazoprevir) and the close up view(right) with hydrogen bonds(orange rectangular), 6 possible hydrogen bonds between nsp13 and the grazoprevir.

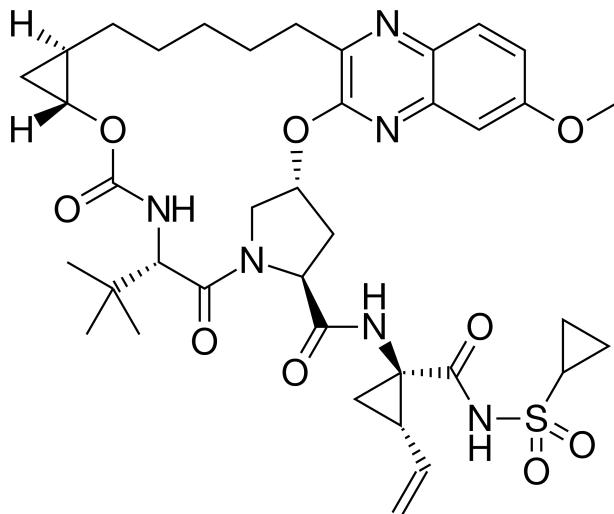


Figure 16: Grazoprevir structure.

Discussion

Evaluation of the result

We created a pipeline with which detected 7 pockets with differing probabilities. We found that the pocket with the highest probability was the ATP pocket, essential to the proper functioning of its ATPase activity. We also found out that the sixth pocket, having the second lowest probability, was involved in the interaction interface between nsp13 and nsp8. We then tested all available ligands from the ECBD database in different conformations on the ATP pocket to find potential inhibitors. The ligand grazoprevir obtained the highest score for protein-ligand interaction and further tests revealed that the molecule could have better interaction with the protein depending on its conformations. We hypothesize that ligands with lower scores are non-competitive ligands since they are shown to have similar scores to the molecule SSYA10-001, known for being a non-competitive ligand. Conversely, grazoprevir and ligands with high scores might be competitive ligands as their scores are similar or higher to the ones obtained from the ATP. However, the test was performed only on a small number of conformations on each ligand, there can be more adaptable ligands for the pocket. And another problem of the result is that it did not concern the decoys of the ligands which are false positive. Even though some ligands show good geometric scores with Convex-PL, it is doubtful that this ligand conformation is feasible in the living cell because there was no consideration of thermodynamic energy terms in the docking and scoring process. The Autodock-vina only considered the electrostatic forces in the inter- and intra-molecular level and the Convex-PL also only considered the geometric conformational scores of the ligands on the pocket. Thus, the tested ligands also need some experimentation in vitro and in vivo to check its reliability. If we use the decoy database to exclude the false positive ligands for the pocket, there might be some real meaningful ligand conformations for the inhibition of the pocket. So, there are some lucks about the grazoprevir suggestion of the pipeline for the best possible ligand for the pocket and the fact that this ligand is already approved ligand for the SARS-CoV-2 helicase in the USA-FDA. Next, we only tested 100 ligands on the interface between nsp13 and nsp8 because of simulation time. Obtained results were not conclusive, results showed that testing the interaction of the ligands on this pocket is possible but we do not have the nsp13-nsp8 interaction score to compare with. We also need to remind ourselves that because of the pocket's low probability, the likelihood for an actual interaction between a ligand and this specific pocket is very low.

NMA analysis

To go further, it is also possible to compare the interaction of nsp13-ligand between the grazoprevir and the ATP in silico. The NMA (Normal Mode Analysis)[13] is already shown that this is effective protein conformation changing simulation method to see the different states of the protein. Some low frequency modes successfully showed the changing states of the protein in many papers. Thus, if we compare the nsp13-grazoprevir with nsp13-ATP using NMA, we might obtain results on conformational difference between these two states. To achieve this, we used NOLB (NOn-Linear rigid Block NMA approach)[14].

As shown on Fig.17 and Fig.18, it is possible to compare the conformational changes

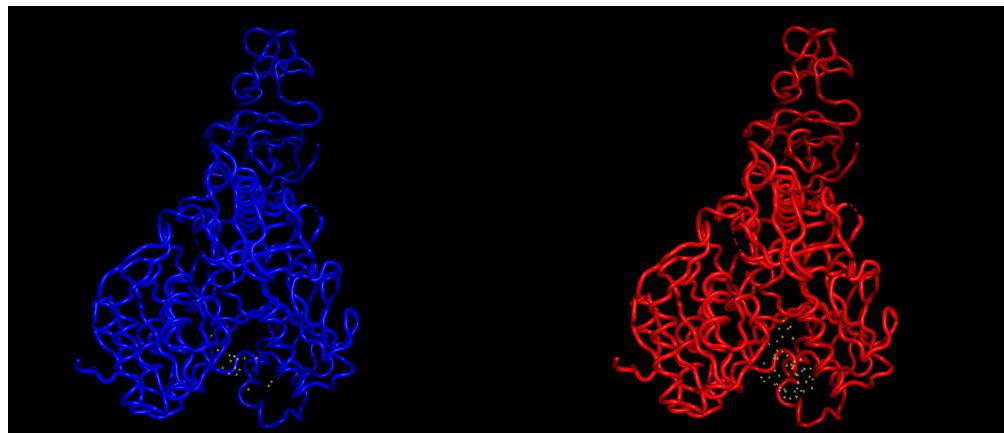


Figure 17: Nsp13-ATP pair(blue) and nsp13-grazoprevir pair(red) of the lowest frequency mode of initial amplitude in NMA. Yellow dots are the ligands.

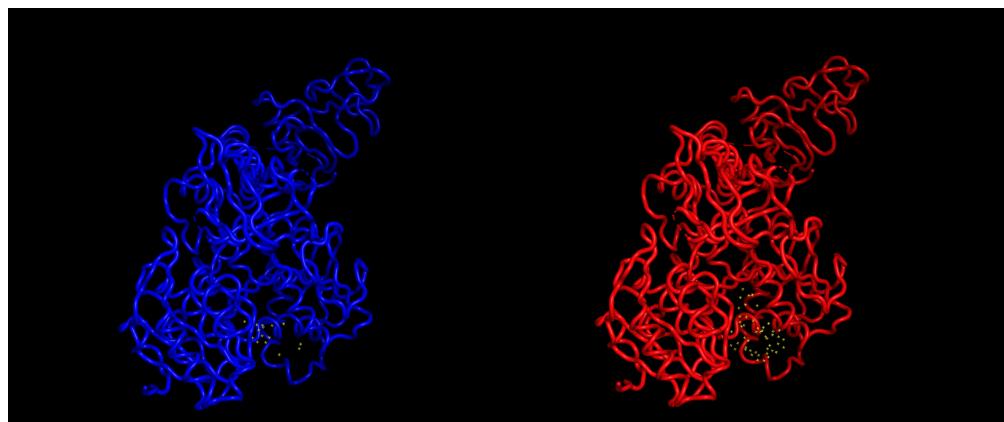


Figure 18: Nsp13-ATP pair(blue) and nsp13-grazoprevir pair(red) of the lowest frequency mode of maximal amplitude in NMA. Yellow dots are the ligands.

of the nsp13 with each ligand and this helps to analyze the impact of the ligands on the protein structure. The first expected observation before running the NMA was that the conformational change of the nsp13-grazoprevir is less vibrant than the nsp13-ATP because grazoprevir has more tendency to interact with nsp13 on the pocket than ATP. It could mean that the grazoprevir pull the nsp13 with more powerful interactive bonds and interrupt the oscillation of nsp13. However, there were no significant changes between nsp13-ATP and nsp13-grazoprevir, one of the possible reason is that the ligands and the nsp13 were linked with hydrogen bonds and the force of hydrogen bonds weren't considered in the NMA method. Even if we couldn't successfully show a meaningful comparison of two models, the NMA analysis is still an interesting tool to simulate the conformational changes of proteins in silico.

References

- [1] *Amber* : <https://ambermd.org>.
- [2] *Autodock Vina* : <https://vina.scripps.edu>.
- [3] Berta, Dénes, et al. “Modelling the Active SARS-CoV-2 Helicase Complex as a Basis for Structure-Based Inhibitor Design”. In: (2021). DOI: <https://doi.org/10.1039/D1SC02775A>.
- [4] Chen, James, et al. “Ensemble Cryo-EM Reveals Conformational States of the Nsp13 Helicase in the SARS-CoV-2 Helicase Replication–Transcription Complex”. In: (2022). DOI: <https://doi.org/10.1038/s41594-022-00734-6>.
- [5] *Convex-PL* : <https://team.inria.fr/nano-d/software/convex-pl>.
- [6] Corona, Angela, et al. “Natural Compounds Inhibit SARS-CoV-2 Nsp13 Unwinding and ATPase Enzyme Activities”. In: (2022). DOI: <https://doi.org/10.1021/acsptsci.1c00253>.
- [7] *Coronavirus Disease (COVID-19) – World Health Organization*. 2023.

- [8] Halma, Matthew T. J., et al. “Therapeutic potential of compounds targeting SARS-CoV-2 helicase”. In: (2022). DOI: <https://doi.org/10.3389/fchem.2022.1062352>.
- [9] Irwin, John J., et al. “ZINC: A Free Tool to Discover Chemistry for Biology . Journal of Chemical Information and Modeling”. In: (2012). DOI: <https://doi.org/10.1021/ci3001277>.
- [10] Ju, Xiaohui, et al. “A Novel Cell Culture System Modeling the SARS-CoV-2 Life Cycle”. In: (2021). DOI: <https://doi.org/10.1371/journal.ppat.1009439..>
- [11] MMFF94 : [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6j490::AID-JCC1j3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6j490::AID-JCC1j3.0.CO;2-P).
- [12] Newman, J. A., Douangamath, A. “Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase”. In: (2021). DOI: [10.1038/s41467-021-25166-6](https://doi.org/10.1038/s41467-021-25166-6).
- [13] NMA : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2279292>.
- [14] NOLB : <https://team.inria.fr/nano-d/software/nolb-normal-modes>.
- [15] p2rank : <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0285-8>.
- [16] Rdkit : <https://www.rdkit.org>.
- [17] Spratt, Austin N., et al. “Coronavirus Helicases: Attractive and Unique Targets of Antiviral Drug-Development and Therapeutic Patents”. In: (2021). DOI: <https://doi.org/10.1080/13543776.2021.1884224>.
- [18] Xiang, Rong, et al. “Recent Advances in Developing Small-Molecule Inhibitors against SARS-CoV-2”. In: (2022). DOI: <https://doi.org/10.1016/j.apsb.2021.06.016>.
- [19] Yan, Liming, et al. “Architecture of a SARS-CoV-2 Mini Replication and Transcription Complex”. In: (2020). DOI: <https://doi.org/10.1038/s41467-020-19770-1>.