MEET-EU Project

Identification of new SARS-CoV-2 NSP13 helicase inhibitors

Sorbonne University Team 4 :

Mouna Ouattara, Fatemeh Kazemi,

Alexis Constensoux, Ekaterina Gaydukova

Supervised by: Prof. Elodie Laine, Prof. Juliana Silva Bernardes

31 January 2024

Contents

1 Introduction

SARS-CoV-2 is the RNA virus responsible for the global COVID-19 pandemic. It is a novel coronavirus that first emerged in late 2019 in Wuhan, China. The virus primarily spreads through respiratory droplets and can cause a range of symptoms, from mild respiratory issues to severe pneumonia, leading to death in some cases.

One of the key components of the SARS-CoV-2 virus is its NSP13 helicase, a protein crucial for viral replication that makes it an excellent antiviral target. The helicase plays a vital role in unwinding the viral RNA, facilitating the replication and transcription processes essential for the virus's life cycle. Inhibiting the NSP13 helicase could be a potential strategy to impede viral replication and, consequently, control the spread of the virus.[1]

This project is composed of multiple steps that include choosing one of the resolved conformation of NSP13 helicase, identifying and characterisation of possible binding pockets, searching and preparation of possible known inhibitors, choosing the tool for virtual screening, implementing the pipeline for virtual screening and scoring that might help to find the best ranked NSP13 helicase inhibitors that can effectively target and inhibit the function of the SARS-CoV-2 virus.

2 Materials and Methods

2.1 Pipeline description

Computational tools play a crucial role in drug design by aiding researchers in various stages of the drug discovery process. Pocket detection tools help identify binding pockets on target proteins where potential drug molecules can interact. Tools like RDkit, Open Babel might be applied for ligand preparation and help to optimize the 3D structure of ligands, ensuring they are in a suitable conformation for docking studies. Docking tools such as AutoDock simulate the binding interactions between ligands and target proteins and make predictions about the preferred binding poses. Virtual screening tools, for example, AutoDock Vina are employed to filter large chemical databases and identify potential lead compounds that are likely to bind to a target of interest.

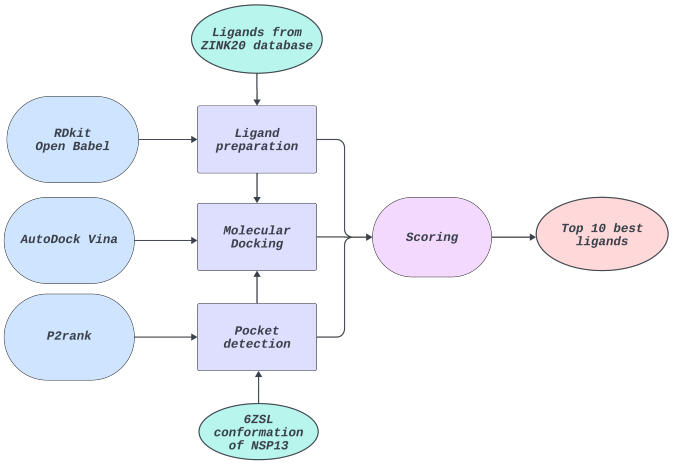

Figure 1: Flowchart of our pipeline

The goal of this project was to create a complete pipeline to identify new potential inhibitors using different packages for consecutive steps of Structure-Based Drug Design. Exploring various tools, P2rank, RDkit, Open Babel and AutoDock Vina were selected for our final pipeline. (Fugure 1)

2.2 NSP13 helicase conformation

The 15 non-structural proteins in SARS-CoV-2 comprise the viral replication and transcription complex which is essential for the coronavirus life cycle.[2] SARS-CoV-2 helicase NSP13 has both ATPase and helicase activity, as it unwinds the RNA helices in an ATP-dependent way. Noting the fact of the high sequence conservation across the coronavirus family, NSP13 might be considered as a potential target for the development of antiviral drugs. The known ATP-binding site of the NSP13 helicase is a promising target for successful inhibition.

The crystal structure of SARS-CoV-2 helicase NSP13 (6ZSL) that was resolved in 2020 has made a big impact in the development of anti-COVID drugs via targeting of NSP13 helicase.[3] SARS-CoV-2 NSP13 helicase have a triangular pyramid shape containing five next domains: the RecA-like domains 1A and 2A, the 2B domain, the zinc-binding domain (ZBD) and the stalk domain (Figure 2). [4]

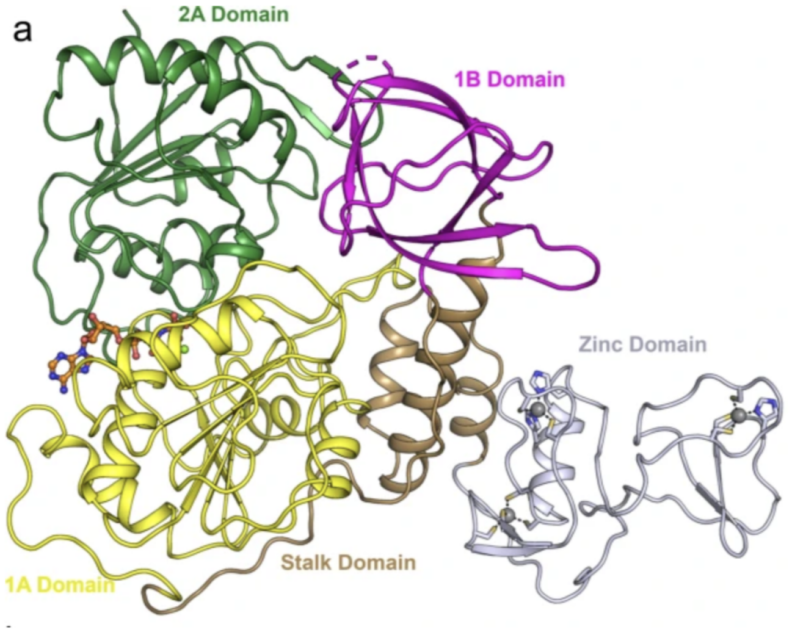

Figure 2: NSP13 helicase domains visualization

We selected 6ZSL resolved crystal structure of the SARS-CoV-2 NSP13 helicase for future identification of binding sites and performing the molecular docking simulations. The PDB structure of the protein was downloaded from the Protein Data Bank site.[5]

2.3 Pocket Detection

With the chosen 6ZSL conformation of NSP13 helicase, the next step in our project was the identification of ligand binding sites. For this task, we have chosen the P2rank. It is an open-source software package that uses protein structure for the searching of ligand binding sites. This tool is based on predicting the ligandicity of local chemical neighborhoods, which are concentrated at points located on the solvent-accessible surface of a protein. P2Rank outperform another tools for ligand binding sites predictions by it speed (it takes less than one second to make a prediction for a single protein). [6]

P2rank has identified 15 potential pockets for 6ZSL conformations. We have chosen the 3D coordinates of the first ranked pocket for our future simulations.

2.4 Database of compounds

The possible inhibitor's molecules were downloaded from the ZINK20 database, a free chemical database of commercially-available compounds for virtual screening. [7] As a result, we have about 5000 molecules that were represented in sdf format.

2.5 Ligands Preparation

After downloading the database of possible ligands, we have used the RDkit tool for the preparation of these molecules for future virtual screening. RDKit is an open-source toolkit for cheminformatics.[8] It is a collection of cheminformatics and machine learning tools that are widely used in the pharmaceutical and chemical industries. For each of the downloaded molecules, we have added explicit hydrogens using Chem.AddHs() and

generated 3D conformations with AllChem.EmbedMolecule(). The modified molecules were also saved in sdf format.

We have used Open Babel after ligands preparation in order to convert our molecules from sdf format to pdbqt format as we needed this type of files for future screening. Open Babel is an open-source chemical toolbox that is widely used in cheminformatics and computational chemistry for tasks such as file format conversion, structure searching, and descriptor calculations.[9]

2.6 Molecular docking and Scoring

Molecular docking is an essential tool in computational biology and drug discovery as it helps predict how small molecules interact with biological macromolecules, providing insights into potential binding sites and the strength of the binding interactions. For virtual screening we have used AutoDock Vina, which is a widely used molecular docking software designed for the prediction of the binding affinity between a small molecule ligand and a target protein.[10] [11]

In order to select the best ligands for chosen binding site of our 6ZSL conformation of NSP13 helicase we have applied an advanced scoring function from AutoDock Vina that could estimate the binding affinity between a ligand and a target protein. The scoring function considers both steric and electrostatic interactions, among other factors, to evaluate the energetics of the ligand-protein complex. The ligands with best scores were chosen as the most probable inhibitors for our 6ZSL conformation founded pocket.

2.7 Pipeline Testing

To validate the efficacy of the implemented pipeline, we conducted preliminary testing using a set of 100 ligands sourced from the ZINK20 database. The output of the pipeline provided a comprehensive list of ligands along with their respective scores, indicative of their

potential possibility of binding with pocket. The results obtained from the pipeline testing on the 100 prepared ligands from the ZINK20 database set demonstrated its capability to successfully identify and score potential inhibitors. The output includes the names of ligands and their corresponding scores, providing valuable insights into their binding affinities.

2.8 High-Performance Computing Deployment

Following the preliminary testing phase, our pipeline was deployed on the High-Performance Computing (HPC) cluster hosted by the Bioinformatics Institute of France. This strategic utilization of the HPC cluster facilitated the access to computational power needed for the execution of molecular docking and scoring procedures across the entirety of the prepared molecules sourced from the ZINK20 database.

3 Results

3.1 Top 10 best-scored ligands

The obtained result, after completing the execution of our pipeline on HPC clusters, was the list of ligands along with their molecular docking scores related to our detected binding pocket in the NSP13 helicase. Subsequently, we sorted this output to identify the top 10 highest-ranked molecules, which now stand as the final results from our pipeline. The top 10 highest-ranked ligands are presented in Table 1

ZINC_NAME	**EOS_NAME**	**Score**
ZINC000008815796	EOS606	8.9
ZINC000008589409	EOS102429	8.7
ZINC000006580945	EOS101914	8.5
ZINC000065232480	EOS1323	8.4
ZINC000043203317	EOS100581	8.2
ZINC000096115282	EOS94	8.2
ZINC000009504660	EOS1047	8.0
ZINC000001547088	EOS100840	8.0
ZINC000011859368	EOS2149	7.9
ZINC000004100613	EOS102423	7.9

Table 1: Table of the the top 10 highest-ranked ligands

3.2 Comparative Analysis of Helicase Inhibitor Search Approaches

This project was a part of the Meet-EU consortium. Our team collaborated with two other teams from Warsaw and Milan Universities, testing different approaches to search for NSP13 helicase inhibitors.

The Warsaw University team explored the application of natural language processing. They employed ProtGPT28, an autoregressive Transformer model designed for generating de novo protein sequences at a high throughput. They explored different docking methods such as HpepDock, CABS-dock, and Alphafold and remarked that the choice of method significantly influences the results of peptide docking simulations. While HpepDock and CABS-dock demonstrated good results for simulations, Alphafold 2 Multimer has some limitations in predictions.

The Milan University team used for Docking and Rescoring VEGAZZ and the Rescore+ tool, also they implemented the models for ATP and RNA binding pockets using Enrichment Factor Optimizer implemented in VEGAZZ. Finally they tested their models on an openscreen pilot library database of compounds and selected best ranked molecules.

Finally, we compared the obtained best-scored ligands by our pipeline and Milan University team's approach and found that we had no matches in these sets of molecules.

4 Discussion

Our structure-based drug design pipeline integrated P2Rank for pocket detection, RDKit for ligand preparation, and AutoDock Vina for molecular docking The collaboration aimed to explore potential binding sites and interactions between ligands and target proteins, essential for drug discovery. Finally delivered best-scored molecules that might be considered as potential inhibitors for NSP13 helicase, but these molecules were not previously described at the literature as potential inhibitors against COVID-19 virus, so it is difficult to validate the efficiency of the obtained results. During the comparison with another team's approach, the differences in predicted inhibitors emphasized the sensitivity of structure-based drug design outcomes to methodological choices. This divergence underscores the need for a nuanced understanding of each tool's strengths and limitations and the impact of different methodologies on drug design outcomes. Future collaborative efforts and exploration of various computational tools might increase the robustness and applicability of the obtained results. Integrating experimental validation with computational predictions remains key to drug design research, bridging the gap between in silico predictions and real-world efficacy.

5 Acknowledgements

We express our gratitude to Professors Elodie Laine and Juliana Silva Bernardes from the Laboratory of Computational and Quantitative Biology at Sorbonne University for their kind guidance during our new experience in international project management and Structure-Based drug design.

We also extend our appreciation to the collaborate teams from Warsaw and Milano Universities for our communication and discussions regarding the results.

Special thanks to the French Institute of Bioinformatics (IFB) for providing the opportunity to utilize the computational resources of the HPC cluster.

Furthermore, we acknowledge the Meet-EU 2023/2024 organization committee team and the 4EU+ consortium for offering us this brilliant opportunity to engage with and contribute to modern and exciting scientific problem.

References

[1] Austin Spratt, Fabio Gallazzi, Thomas Quinn, Chris Lorson, Anders Sönnerborg, and Kamal Singh. Coronavirus helicases: attractive and unique targets of antiviral drug-development and therapeutic patents. *Expert opinion on therapeutic patents*, 31, 02 2021.

[2] Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. Coronavirus biology and replication: implications for sars-cov-2. *Nature Reviews Microbiology*, 19(3):155–170, 2021.

[3] Joseph A Newman, Alice Douangamath, Setayesh Yadzani, Yuliana Yosaatmadja, Antony Aimon, José Brandão-Neto, Louise Dunnett, Tyler Gorrie-Stone, Rachael Skyner, Daren Fearon, et al. Structure, mechanism and crystallographic fragment screening of the sars-cov-2 nsp13 helicase. *Nature communications*, 12(1):4848, 2021.

[4] RP Vivek-Ananth, Sankaran Krishnaswamy, and Areejit Samal. Potential phytochemical inhibitors of sars-cov-2 helicase nsp13: A molecular docking and dynamic simulation study. *Molecular diversity*, 26(1):429–442, 2022.

[5] Newman et al. 6zsl crystal structure of the sars-cov-2 helicase. `https://www.rcsb.org/structure/6zsl`, 2020.

[6] Radoslav Krivák and David Hoksza. P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of cheminformatics*, 10:1–12, 2018.

[7] John J Irwin, Khanh G Tang, Jennifer Young, Chinzorig Dandarchuluun, Benjamin R Wong, Munkhzul Khurelbaatar, Yurii S Moroz, John Mayfield, and Roger A Sayle. Zinc20—a free ultralarge-scale chemical database for ligand discovery. *Journal of chemical information and modeling*, 60(12):6065–6073, 2020.

[8] Rdkit : Open-source cheminformatics. https://www.rdkit.org.

[9] Open babel : chemical toolbox designed to speak the many languages of chemical data. http://openbabel.org/index.html.

[10] Jerome Eberhardt, Diogo Santos-Martins, Andreas F Tillack, and Stefano Forli. Autodock vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, 61(8):3891–3898, 2021.

[11] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.