EUROPEAN BIOINFORMATICS MASTERS NETWORK
Meet-EU Course 2022/2023

Prediction of potential inhibitors for the Sars-CoV-2 Helicase (NSP13) by virtual screening and MD simulations

Final report

By **Team 2** from **Ruprecht-Karls-University of Heidelberg**:
Elizaveta Chernova,
Faculty of Biosciences, Master of Science Biochemistry
Valeriia Dragan,
Faculty of Engineering Sciences, Master of Science Molecular Biotechnology

Supervised by PD Dr. Carl Herrmann,
Dr. Gurdeep Singh, Magdalena Georgieva
Health Data Science Unit - BioQuant & Medical Faculty Heidelberg

Submitted February 06, 2023

Introduction

SARS-CoV-2 virus, the causative agent of the COVID-19 pandemic, has caused widespread illness and death worldwide. One of the SARS-CoV-2 proteins responsible for viral replication is the non-structural protein 13 (NSP13). NSP13 is a helicase from the 1B superfamily and plays a crucial role in viral replication and pathogenesis [1]. Due to its high sequence conservation, NSP13 is an attractive target for drug development [2]. In this work, virtual screening of the European Chemical Biology Database (ECBD) was employed to predict potential NSP13 inhibitors. Molecular dynamics simulations provided in-depth analysis of molecular interactions between NSP13 and the proposed inhibitors. Results of this study may act as a starting point for anti-viral drug development targeting NSP13.

Previously several small molecule NSP13 inhibitors, such as adamantane-derived structures and RNA-targeting compounds, were described [3]. These molecules showed promising results during *in vitro* testing [4].
Similarly to previous approaches, we decided to conduct virtual screening for the identification of possible inhibitors among molecules present in ECBD [5]. To this end, we used a common docking software, AutoDock Vina [6]. An overview of the complete workflow can be seen in Figure 1. As docking may be computationally intensive, the database was pre-filtered to only include compounds with low toxicity scores predicted by the web tool ADMETlab 2.0 [7]. In order to increase docking accuracy, NSP13 binding pockets were predicted by combining output from Fpocket and P2Rank tools [8,9]. Molecular dynamics (MD) simulation was used to analyze molecular interactions between the NSP13 protein and the proposed inhibitors. Finally, possible influence of known mutations in the sequence of NSP13 on binding of the proposed inhibitors was assessed.

Materials & Methods

NSP13 structure

NSP13 structures with IDs 7NIO and 7NN0 were obtained from the Protein Data Bank (PDB). The structures were pre-processed by MeetEU organisers. The 7NN0 structure is NSP13 bound to an ATP-analog, while 7NIO is the apoenzyme.
These structures were chosen because they best characterise the structural changes of the helicase during its catalytic cycle. The 7NIO structure consisted of two chains, while the 7NN0 consisted of 4 chains. The chains were separated and treated as replicas.

Binding sites identification

The binding pocket prediction was performed using two platforms: Fpocket and P2Rank [8,9]. Fpocket uses geometrical parameters to identify binding pockets. P2Rank is a machine learning-based tool for binding pocket identification. For both tools, the default settings were used.
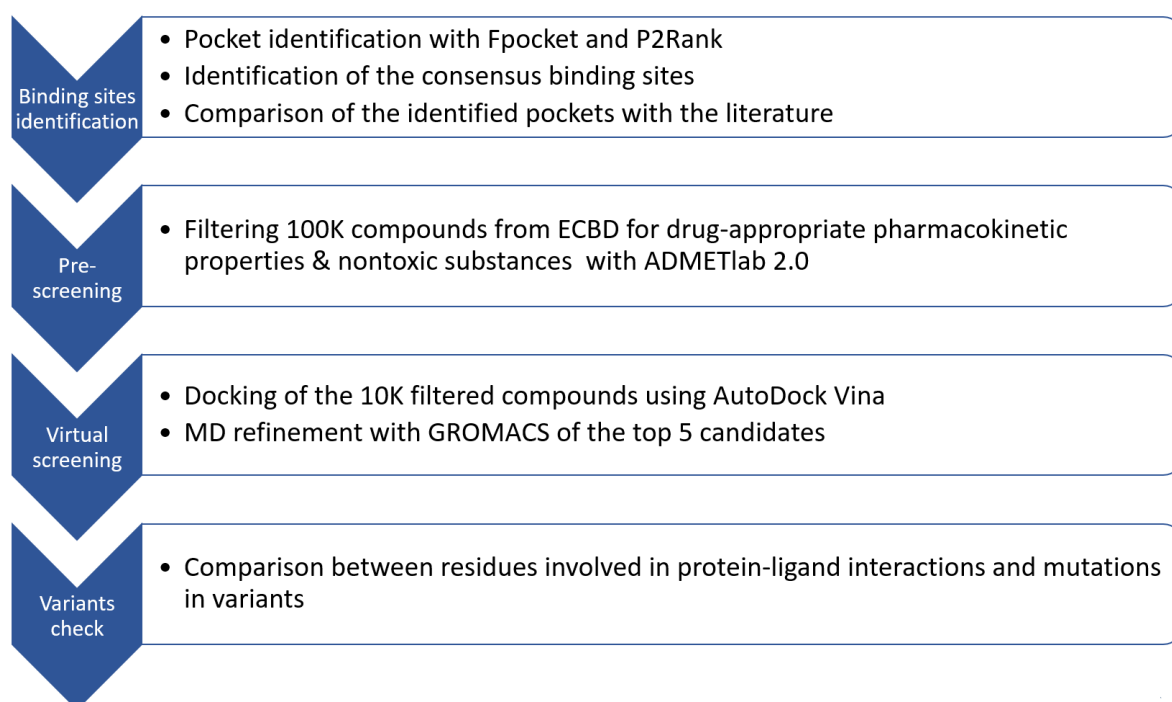
Figure 1. The workflow was performed for both 7NIO and 7NN0 NSP13 structures.

Predictions were done for each chain separately. Amino acids that build up the predicted binding pockets were retrieved from both tools. P2Rank yields residue numbers directly. For Fpocket numbers of residues in a radius of 6 Å from the centres of the so-called alpha-spheres were collected. Overlaps between the residues were calculated. The pockets predicted for different chains were merged. As both Fpocket and P2Rank provided scoring for the putative binding scoring, the ranking of the merged pockets was defined by comparing the mean scores of the initial pockets.

Screening Database

A total of 101,995 compounds were retrieved from the European Chemical Biology Database [10] in December 2022. The drugs were downloaded in SMILES format.
Compounds were filtered based on their ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties. Sixteen different toxicities were predicted by ADMETlab 2.0 web tool [7]. Compliance with a set of heuristic rules (Lipinski Rule, Pfizer Rule, GSK Rule and Golden Triangle Rule) was also evaluated by ADMETlab 2.0. All compounds that violated at least one rule or had more than 6 high toxicity predictions were eliminated. After filtering, a total of 10,453 molecules passed the filter and were used for docking.

Molecular Docking

High-throughput docking of the ligands was performed using the AutoDock Vina (version 1.2.3) program. AutoDock Vina utilises a conformational search strategy based on gradients and an empirical scoring function [6]. It also makes use of parallel computing.
Vina force field with default parameters was used. For both of the structures (7NIO and 7NN0) the A chain and the best-ranked pocket were used. For 7NN0, the centre of mass of

the pocket was used to centre the binding box. The size of the box was 25 x 25 x 25 Å. For 7NIO, the best-ranked pocket had a curved shape, therefore the box was adjusted to be centred at (-25.229235, 25.126171, -28.270075). The size of the box was 30 x 30 x 30 Å. The expected best score for these settings is described as -13 kcal/mol [11].

To avoid writing the ligand structures on the disk in PDB format, Meeko and RDkit python packages were used to make the docking possible directly from ligand SMILES strings. After the initial docking with an exhaustiveness of 24, the top compounds with scores below -11 kcal/mol were selected. The docking was repeated with the higher exhaustiveness of 32.
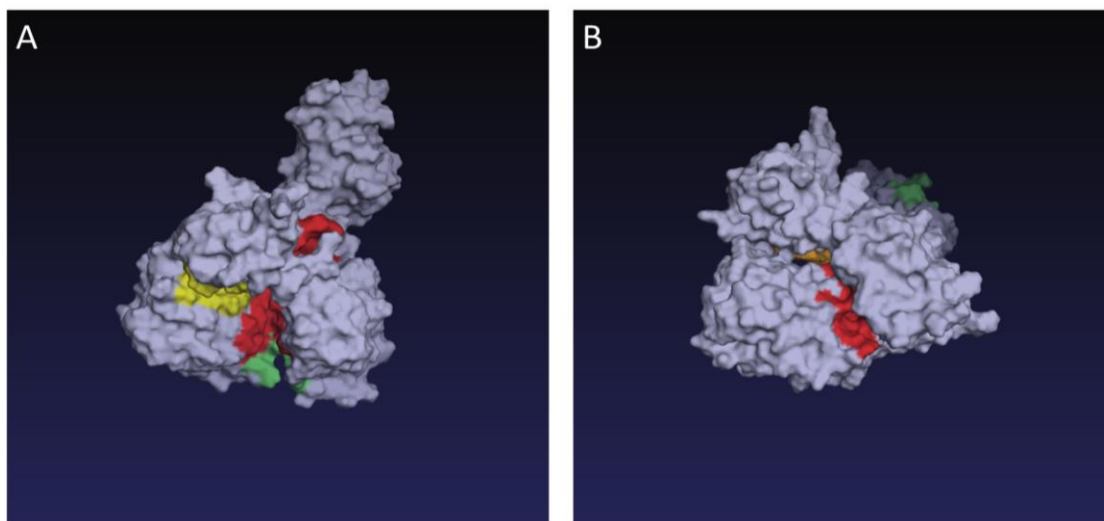

Figure 2. Merged pockets. Pockets with the highest rank (used in later analysis) are shown in red. A. 7NIO B. 7NN0.

Molecular Dynamics simulations

The five top-ranked putative inhibitors identified by docking were used to perform MD simulations. Docked complexes of ligands and the helicase were used as the starting point. The simulation software GROMACS (version 2022.4) was used with CHARMM36 (version July 2022) force field [12,13]. To prepare ligand topology the web tool CGenFF (version 4.6) was used [14,15]. The TIP3P CHARMM-modified water model was used [16]. The simulations were performed in a dodecahedron box with periodic boundary conditions.

The Na+ or Cl− counter ions were added to obtain a neutralised system with physiological ion concentrations. After the energy minimization and equilibration step, 10 ns MD simulation was performed under the constant temperature of 300 K. The change in distance between the centres of masses of the ligand and the protein over time, as well as the root-mean-square deviation of the ligand were obtained from the simulation. Moreover, hydrogen bonds between the ligand and the protein could be detected using tools provided by GROMACS.

Variant analysis

Known SARS-CoV-2 mutations were retrieved from the dynamically updated NCBI database (accessed on 15.12.2022)[17]. Non-synonymous mutations overlapping with binding pockets

were detected. They were then compared to residues involved in hydrogen bond formation between the ligand and the protein.

Python version and packages

The Python version used was 3.8.15. Packages used can be found in the .yml environment file on GitHub.

Results

Binding pocket identification

To reduce the search space for the molecular docking algorithms, we decided to first predict the binding pocket. After merging the results of the Fpocket and P2Rank tools, three binding pockets for the 7NN0 structure and five pockets for the 7NIO structure were obtained. Residue numbers of amino acids that build up pockets can be found on GitHub. The location of the pockets can be seen in Figure 2.

We decided to use the binding pockets with the highest rank for both structures. The Jaccard index (intersection over union) of the residues that constitute these pockets was 33%. The pocket from the 7NIO structure was larger than the pocket from the 7NN0 structure. Therefore, 90% of residues in the 7NN0 pocket overlap with the 7NIO pocket, while only 34% of residues from the 7NIO pocket overlap with the 7NN0 pocket.
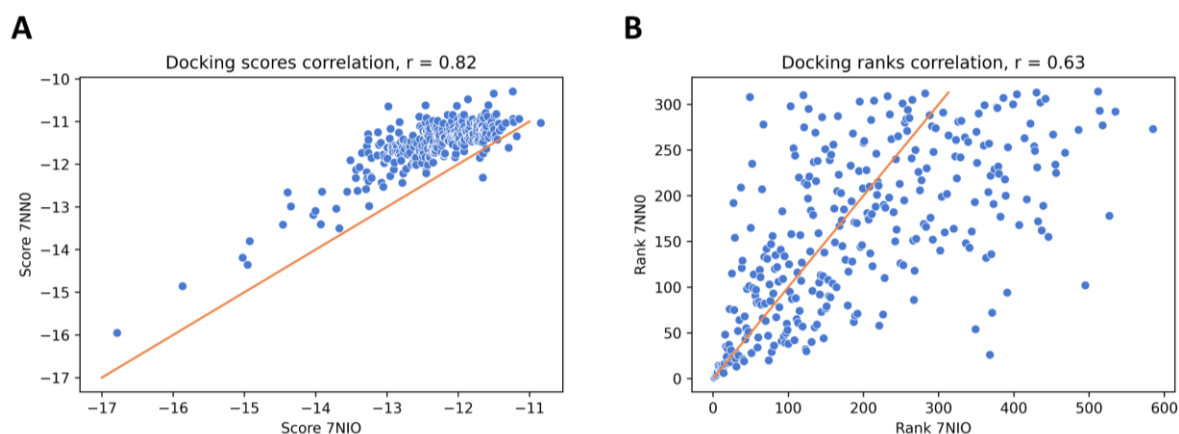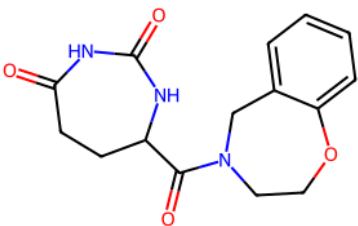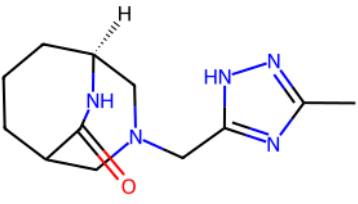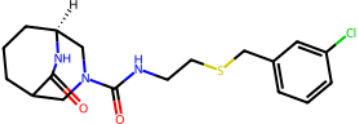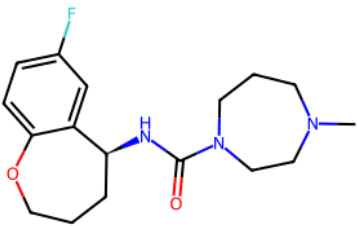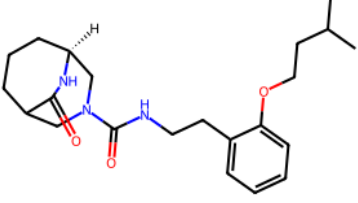

Figure 3. A. Correlation between docking scores predicted for the structure 7NIO and the structure 7NN0. B. Correlation between ranks predicted for the structure 7NIO and the structure 7NN0. Blue dots represent single molecules. Shown in orange is the identity line. r = Pearson correlation coefficient.

Molecular docking results

Due to limited time and computational resources, we performed docking only on 10,453 molecules from the ECBD database with the lowest toxicity predictions. After initial docking, molecules with scores above -11 kcal/mol were discarded. This resulted in a set of 593 molecules for the 7NIO and 314 molecules for the 7NN0 structure. All the molecules from the 7NN0 set were also present in the 7NIO set. After the refinement docking, the same

molecules for both structures were identified as the top five hits, but with different ranking (Table 1). For all five molecules, ADMETlab web tool predicted low toxicity. Lipinski, Golden Triangle, and GSK rules were positive for all five molecules. Pfizer rule was positive for all molecules except ligand 5. None of the molecules contained PAINS (pan-assay interference compounds).

Table 1. Top five best-scored ligands after refinement docking.

Ligand	Structure	Rank 7NIO	Rank 7NN0	Score 7NIO	Score 7NN0
1		1	1	-16.786	-15.953
2		2	2	-15.867	-14.855
3		3	4	-15.026	-14.191
4		4	3	-14.954	-14.358
5		5	5	-14.926	-13.802

The Tanimoto coefficient between all possible pairs of molecules was computed to determine whether the docking software scores similar molecules comparably (Figure 4)[18].

To assess whether the docking results are consistent, the correlation of scores and ranks between the 7NIO and 7NN0 structures was analysed (Figure 3). The Pearson correlation coefficient was 0.82 for the docking scores and 0.63 for the ranks.

The next question was whether molecules with similar structures are enriched in a set of 300 molecules with the highest docking score. The distribution of Tanimoto coefficients of all possible pairs of molecules among the 300 highest ranked for each structure was compared to that of random molecules from the initial dataset of 10,453 molecules (Figure 4).
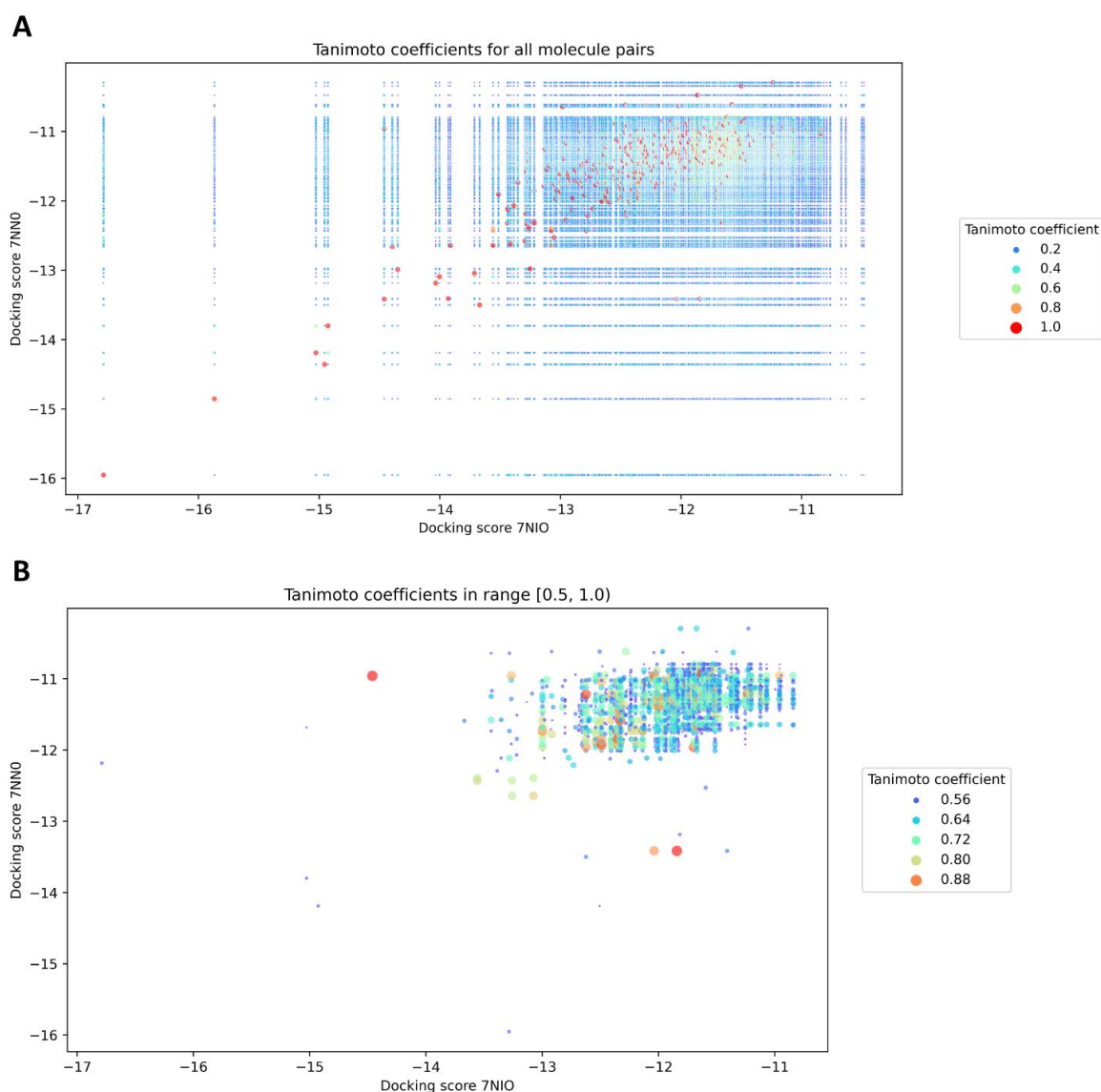

Figure 4. A. Tanimoto coefficient of all possible pairs of molecules (including the comparison of molecule to itself). Each point represents a pair of molecules with docking scores as seen in the axis. B. Only pairs that have the Tanimoto coefficient of at least 50% and do not consist of identical molecules (Tanimoto coefficient of 100%) are shown.
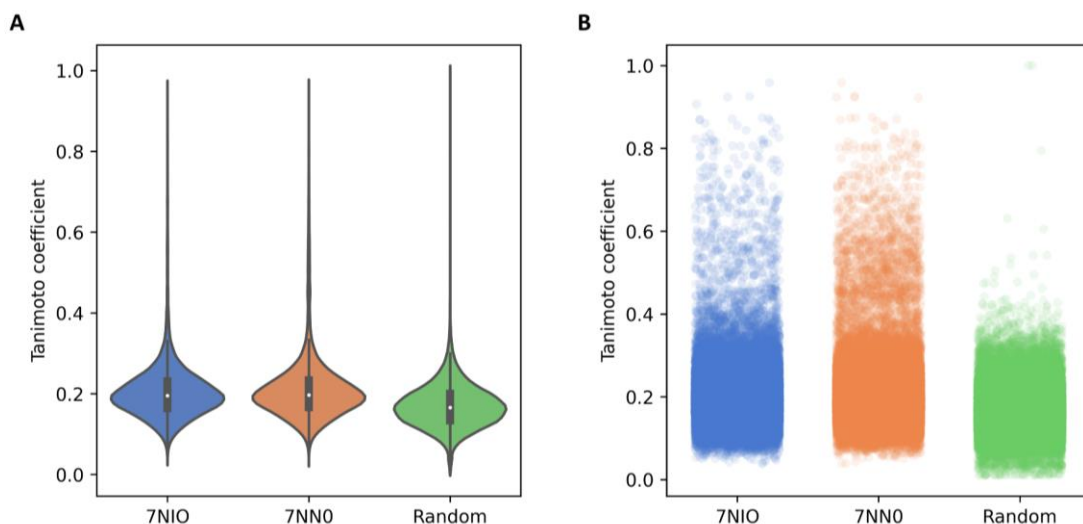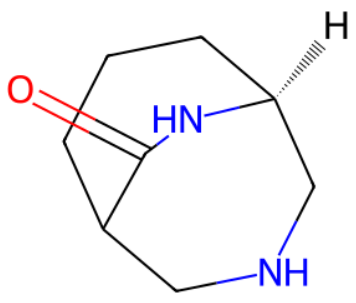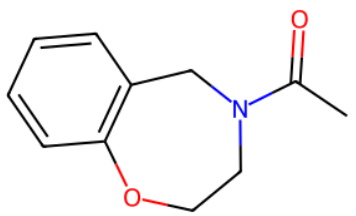
Figure 5. Distribution of Tanimoto coefficient values inside different groups. All possible pairs inside a group were analysed. Pairs of two identical molecules were eliminated. A. Violin plot. B. Each point corresponds to a pair of molecules.

One common substructure was identified in three from the top five molecules. Additionally, a substructure from ligand 1 similar to a substructure in ligand 4 was analysed (Table 2).

Table 2. Substructures identified in the five top ranked molecules after docking.

Substructure	Chemical structure	Occurrence
1		Ligands 2, 3 and 5
2		Ligand 1, similar structure in ligand 4

To determine whether these substructures may contribute to better binding to the target, the docking score distribution between the molecules containing the substructures and the other molecules was compared (Figure 6).
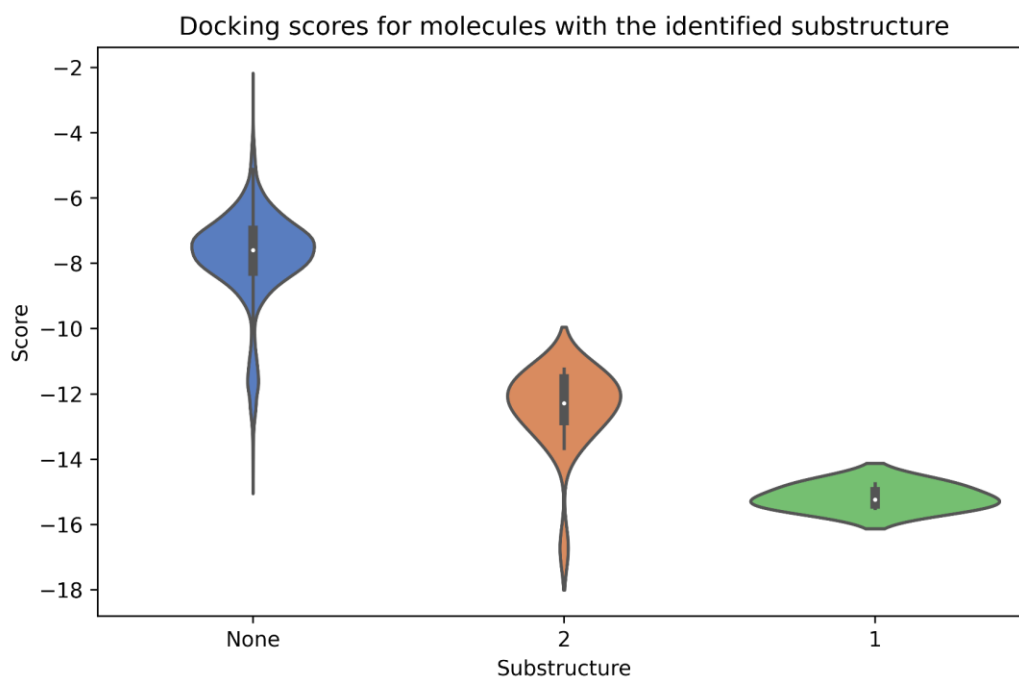
Figure 6. Distribution of docking scores from the screening docking step for molecules that either include one of the identified substructures or not.

In the initial dataset used for docking, there were only 3 molecules with the substructure 1 and all of them were ranked in the top five hits for both structures. From 22 structures containing substructure 2, 16 had docking scores below -11 and were also considered in the refinement docking step.
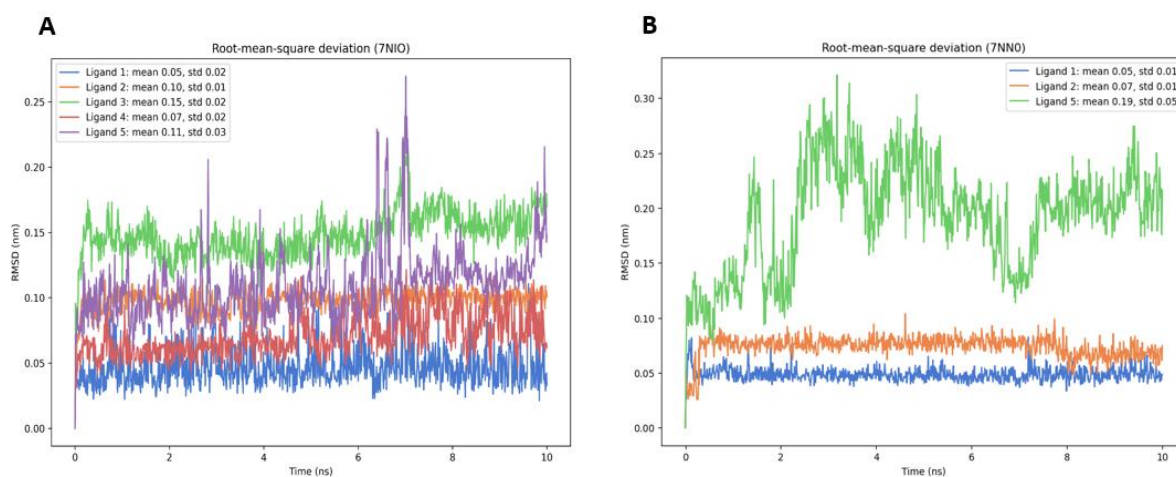

Figure 7. RMSD of the top molecules over time. A. 7NIO B. 7NN0

Molecular dynamics simulation results

For further description of predicted molecule binding, 10 ns molecular dynamics simulations were performed on the molecules with the highest docking scores. In Figure 7 you can see the root-mean-square deviation of the molecules over time.

Hydrogen bonds were detected based on geometric criteria using GROMACS software. Only bonds that persisted longer than 20% of the simulation time (2 ns) were considered. Atoms that participated in the hydrogen bonds are listed in Table 3.

Atoms and corresponding residues participating in hydrogen bonds can be found in Table 3.

Table 3. Residues and atom names of atoms involved in hydrogen bond formation for 7NIO.

Structure / Ligand	Bond number	Donor	Acceptor
7NIO / Ligand 1	1	LEU 418, N	Ligand, O3
	2	Ligand, N3	GLY 416, O
7NIO / Ligand 2	1	ARG 410, N	Ligand, N2
7NIO / Ligand 3	1	Ligand, N3	TYR 181, OH
7NIO / Ligand 4	1	ARG 410, N	Ligand, O1
7NIO / Ligand 5	1	ARG 410, N	Ligand, O2
	2	LEU 418, N	Ligand, O3

Virus variant analysis

To evaluate the possible influence of known genome mutations of SARS-CoV2 on the binding of the proposed inhibitors, mutations occurring in the identified binding pockets were retrieved from the NCBI database (Table 4). None of these mutations included residues involved in hydrogen bonds during molecular dynamics simulation.

Table 4. Identified mutations in the considered binding pockets.

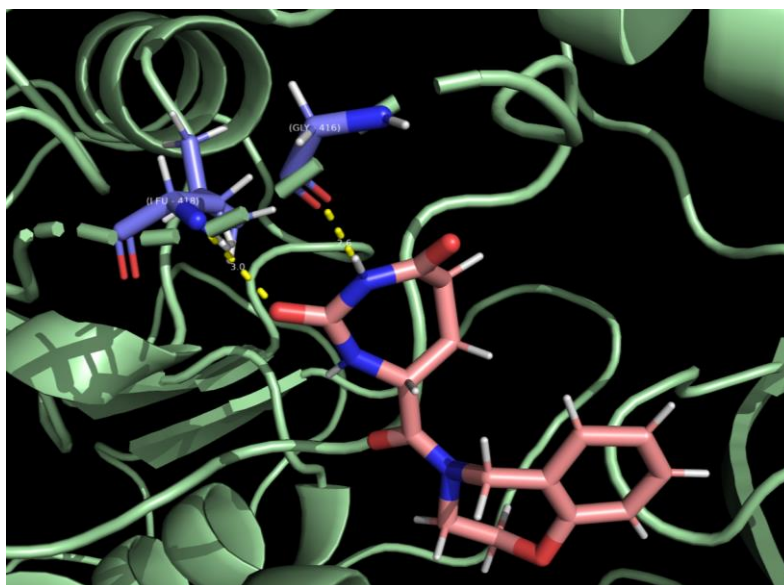| Structure attribution of the binding pocket	Mutations
7NIO	H290Y K460R M233I M249I
7NN0	H290Y

Figure 8. Hydrogen bonds (yellow) between the ligand 1 (pink) and the protein (structure 7NIO). Interacting amino acids are shown in light blue. Oxygen is shown in red, nitrogen in blue and hydrogen in white.

Comparison of results with the Prague Team 2

Prague Team 2 used a selection of molecules from ZINC and PubChem databases and the PLANTS docking software. Therefore, direct comparison of predicted molecules or docking scores was not possible. In Figure 9 graphical representation of Tanimoto coefficient for all possible pairs of molecules used by us and by Prague Team 2 is shown. Ranks for our molecules were recalculated based on the average score for the 7NIO and the 7NN0 structure.
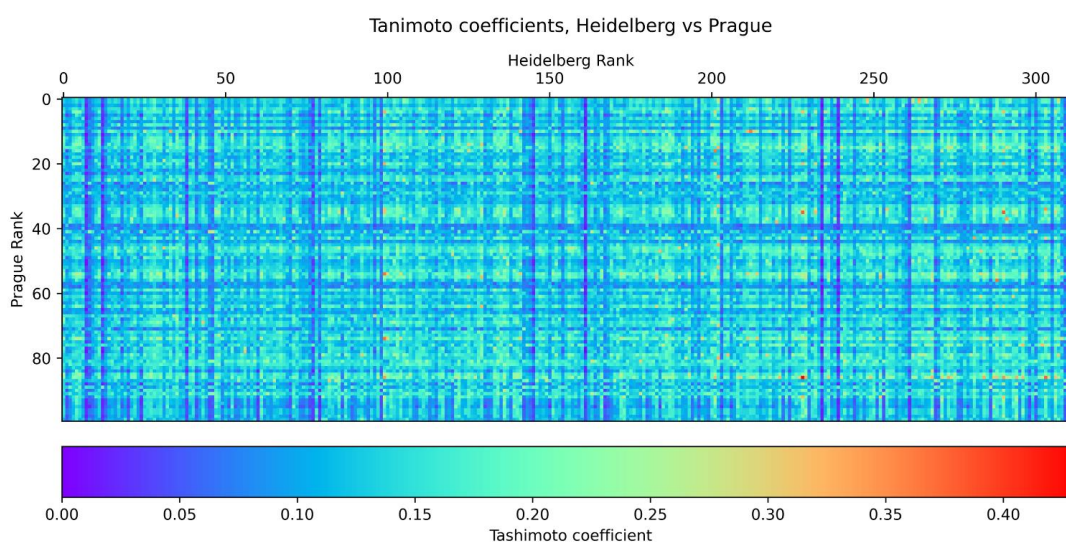

Figure 9 Tanimoto coefficients of all possible pairs of molecules between Prague and Heidelberg datasets. Each pixel represents a pair of molecules. The horizontal coordinate represents the rank of the molecule from the Heidelberg Team, while the vertical coordinate shows the rank of the molecule from the Prague Team.

None of the molecules used by Prague Team 2 contained either the earlier described substructure 1 or substructure 2.

Discussion

Identification of the binding sites was quite accurate, as it recovered the ATP binding pocket as the most probable [2]. Moreover, the results were similar to the previous efforts in computational detection of the NSP13 binding pockets [19]. As we restrained virtual screening to the ATP binding pocket, it would be interesting to look at the allosteric binding sites in the future.

Docking showed consistent results for different NSP13 structures, as it yielded identical compounds as the top five candidates and showed high correlation between results for both the 7NIO and the 7NN0 structures. In the future, it would be interesting to perform docking on a larger number of molecules with higher exhaustiveness to get more precise results. Analysis of docking results showed that highly scored molecules are more similar than random molecules, which was expected. Moreover, two substructures contributing to the binding of the ligands were detected. Further on, more ligands with identified structures could be analysed to find a better drug candidate.

Molecular dynamics has shown that the ranking of potential ligands based on ligand RMSD does not reproduce the ranking based on docking scores. These results are in agreement with the previous studies [20]. This suggests that for more precise results, MD simulations of more candidates should be performed. However, due to the high computational cost, the number of candidates will remain limited, and the optimal number of candidates for the test remains to be refined.

As the MD results showed, the residues involved in the protein-ligand interaction do not overlap with the mutated residues in the variants. This suggests that the inhibitor candidates established in this study would be potential universal inhibitors for the SARS-CoV-2 strains considered above.

The pipeline used in this study is subject to further validation. One of the options would be to compare the effectiveness of the pipeline with the one from the Prague Team 2.

Overall, the aim of the work to detect small molecule NSP13 inhibitors based on the protein structure binding pockets was successful. Application of the pipeline on a larger library would have been necessary before *in vitro* validation.

References

1. Chen, J. *et al.* Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **182**, 1560-1573.e13 (2020).

2. Newman, J. A. *et al.* Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat. Commun.* **12**, 4848 (2021).

3. Perez-Lemus, G. R., Menéndez, C. A., Alvarado, W., Byléhn, F. & de Pablo, J. J. Toward wide-spectrum antivirals against coronaviruses: Molecular characterization of SARS-CoV-2 NSP13 helicase inhibitors. *Sci. Adv.* **8**, eabj4526 (2022).

4. Wang, Z. *et al.* On the mechanisms of bananin activity against severe acute respiratory syndrome coronavirus. *Febs J.* **278**, 383–389 (2011).

5. Piplani, S., Singh, P., Winkler, D. A. & Petrovsky, N. Potential COVID-19 Therapies from Computational Repurposing of Drugs and Natural Products against the SARS-CoV-2 Helicase. *Int. J. Mol. Sci.* **23**, 7704 (2022).

6. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

7. Xiong, G. *et al.* ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* **49**, W5–W14 (2021).

8. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).

9. Krivák, R. & Hoksza, D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminformatics* **10**, 39 (2018).

10. ECBD Home. https://ecbd.eu/.

11. Basic docking — Autodock Vina 1.2.0 documentation. https://autodock-vina.readthedocs.io/en/latest/docking_basic.html.

12.   Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).

13.   MacKerell Jr., A. D. *et al.* CHARMM: The Energy Function and Its Parameterization. in *Encyclopedia of Computational Chemistry* (John Wiley & Sons, Ltd, 2002). doi:10.1002/0470845015.cfa007.

14.   Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).

15.   Vanommeslaeghe, K. & MacKerell, A. D. Jr. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).

16.   MacKerell, A. D. Jr. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

17.   SARS-CoV-2 Variants Overview. https://www.ncbi.nlm.nih.gov/activ.

18.   Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminformatics* **7**, 20 (2015).

19.   Ricci, F., Gitto, R., Pitasi, G. & De Luca, L. In Silico Insights towards the Identification of SARS-CoV-2 NSP13 Helicase Druggable Pockets. *Biomolecules* **12**, 482 (2022).

20.   Guterres, H. & Im, W. Improving Protein-Ligand Docking Results with High-Throughput Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **60**, 2189–2198 (2020).