Structure-based screening in search of SARS-CoV-2 Nsp13 helicase inhibitors

Anna Szymik, Bruno Puczko-Szymański, Jagoda Trzeciak, Kacper Chętkowski, Marta Korpacz, Oleksandra Shcherbyna

February 2024

1. Introduction

The main aim of the project was to identify candidates for SARS-CoV-2 Nsp13 inhibitors, using a set of resolved enzymatic structures as a starting point.

We chose to use a so-called classical approach for the drug discovery, namely Structure-based High-Throughput Virtual Screening (HTVS), which enables rapid evaluation of binding of millions of compounds. The project workflow with the key points is presented in Fig. 1. and described in more detail in the Materials and Methods section.

During the project we were paired with the two other teams, one from Heidelberg University and the other from Sorbonne University (S5).

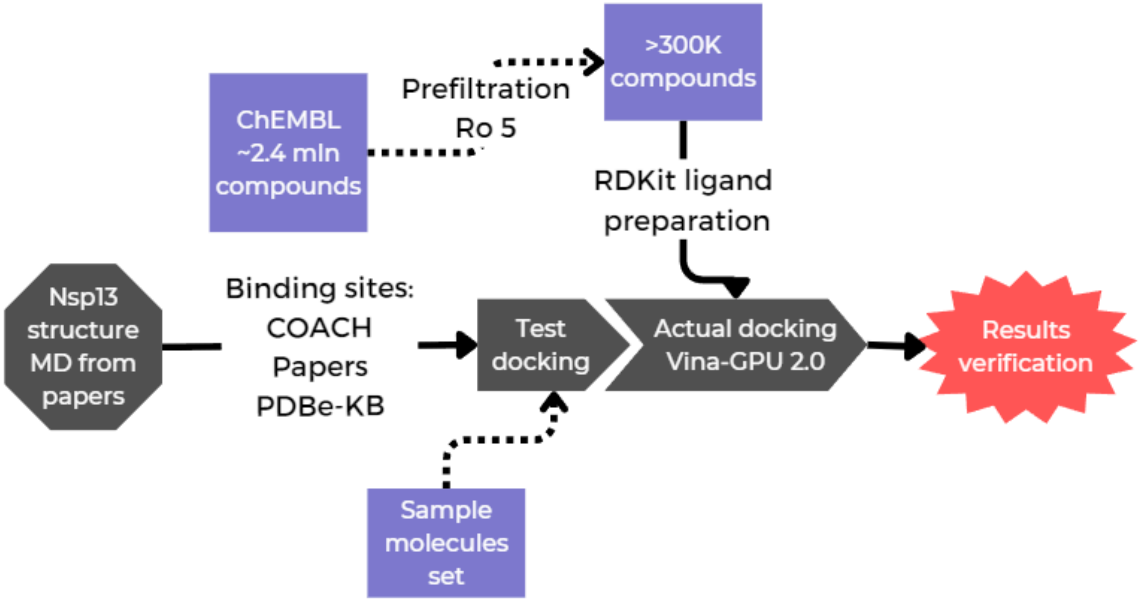

Figure 1. The project workflow with the key points.

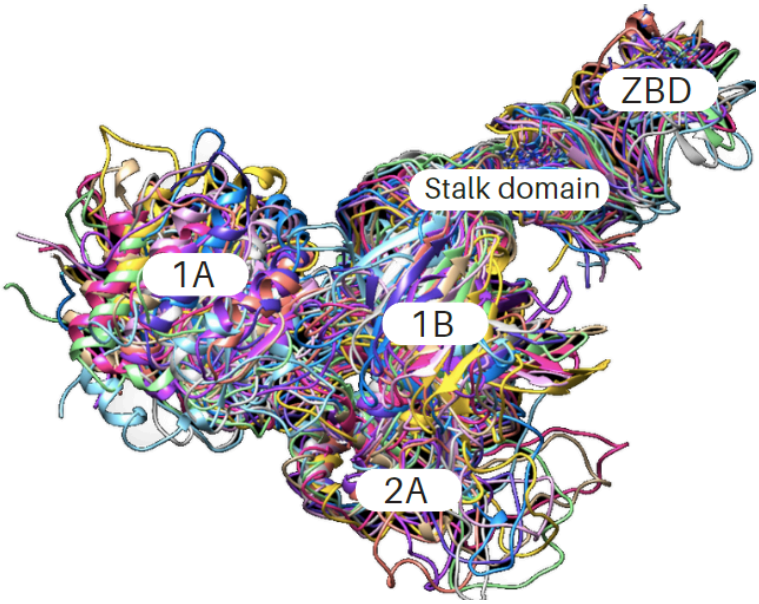

Figure 2. Nsp13 MD structures superposition visualization, based on MD data obtained from Raubenolt et al.[1]

2. Materials and Methods

2.1. Potential binding sites identification

SARS-CoV-2 Nsp13 helicase is organized into 5 distinct domains. The ATP nucleotide binding site is situated between the domains 1A and 2A, with specific contacts to the nucleotide provided by conserved helicase motifs I, II, and III in the 1A domain and IV, V, and VI in the 2A domain.[2] As the ATP nucleotide processing is crucial for the helicase function, we decided to focus on the area of its binding.

A binding pocket was selected by rationally integrating binding residues data from COACH tool for the site prediction[3], PDBe-KB database[4] and residues identified in previous studies [1,2]. These are summarized in Table 1. Residues present in the final docking box (Fig.2) are depicted in Table 2. and are all part of the Nsp13 ATP binding site.

a)

Binding residue	289	290	320	374	375	404	442	443	540	567
Number of ligands	6	12	7	2	3	3	12	5	2	4

b)

Binding residue	256	257	278	279	280	281	282	283	284	285	396	434	435	532	559
Number of ligands	X	X	X	X	X	1	1	2	4	11	X	2	X	1	X

c)

Binding residue	129	132	285	286	287	290	442
Number of ligands	9	9	11	13	15	12	12

Table 1. Number of the ligands assigned to the binding residues of Nsp13 according to the research papers [1,2] **a)**, COACH site prediction tool **b)**, and the PDBe-KB database **c)**. Only residues with the highest number of assigned ligands are presented.

Binding residue	Structural motif
Q281–K288	phosphate binding motif I
Q404	phosphate binding motif III
R567	phosphate binding motif VI

Table 2. The binding residues presented in the docking box.

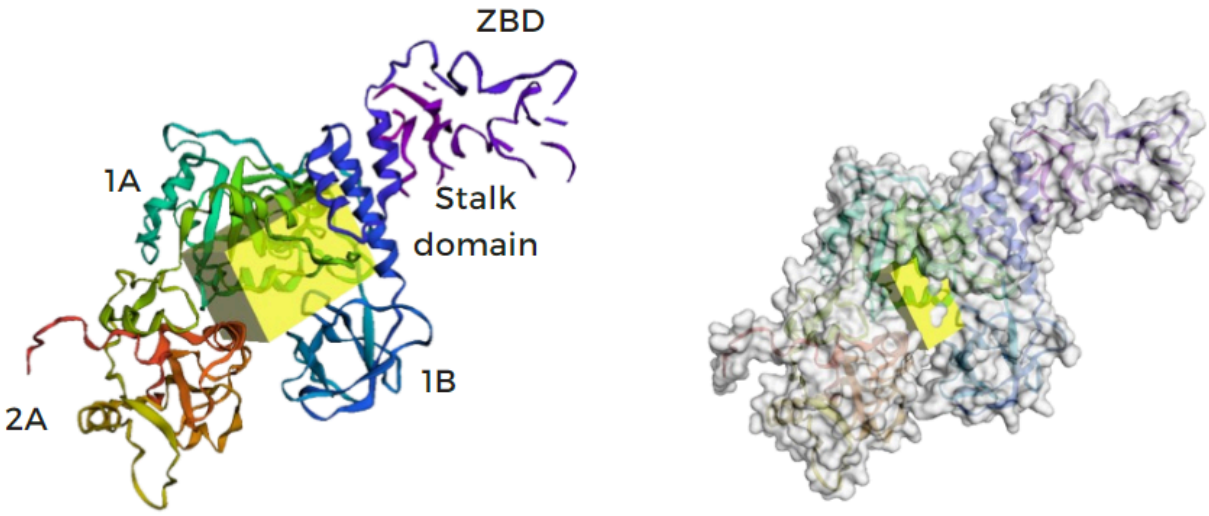

Figure 3. Nsp13 docking box visualized by Webina 1.0.5[5].

2.2. Ligand prefiltration

Compounds from the ChEMBL[6,7], a manually curated database of bioactive molecules with drug-like properties, with an initial amount around 2,4 mln were prefiltered in terms of Lipinski's rule of five (Ro 5)[8] with the criteria depicted in Table 3.

Molecular mass	< 500 Da
Octanol/water partition coefficient (CLogP)	< 5
Number of hydrogen bond donors	< 5
Number of hydrogen bond acceptors	< 10

Table 3. Lipinski's rule of five criteria.

2.3. Virtual screening and docking

Prefiltrated ligands and protein structures were preprocessed using the open-source chemoinformatics tool RDKit[9], a screening test was then performed using a set of Nsp13 inhibitors known from previous studies[10] with an aspirin as a negative control (Fig. 4).

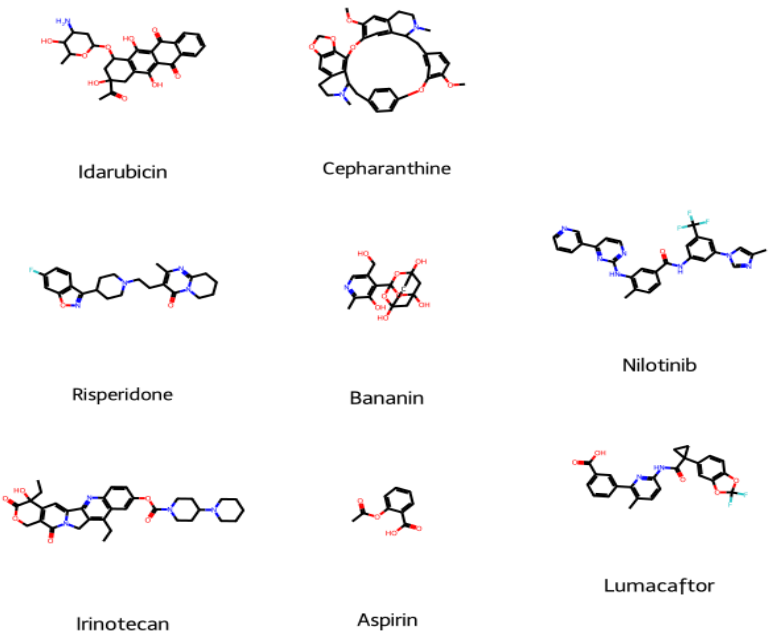

Figure 4. A set of ligands for the test screening.

After testing, virtual screening and rigid docking was performed on a representative protein structure using Vina-GPU 2.0[11,12] and the process was accelerated using CUDA Toolkit 11.5[13].

Next, ligands with the free energy of binding less than -10 kcal/mol were analyzed in terms of potential toxicity and synthetic accessibility using eToxPred tool[14].

3. Results

3.1. High-Throughput Virtual Screening

346,885 ligands obtained from the prefiltration were used for the docking. Distribution of the free energy of binding for all docked compounds is presented in Figure 5.

The result distribution is close to normal, slightly right-skewed; therefore it does not give rise to any concerns[15]. Certain outcomes featuring low binding free energy were obtained and subjected to a more in-depth analysis.

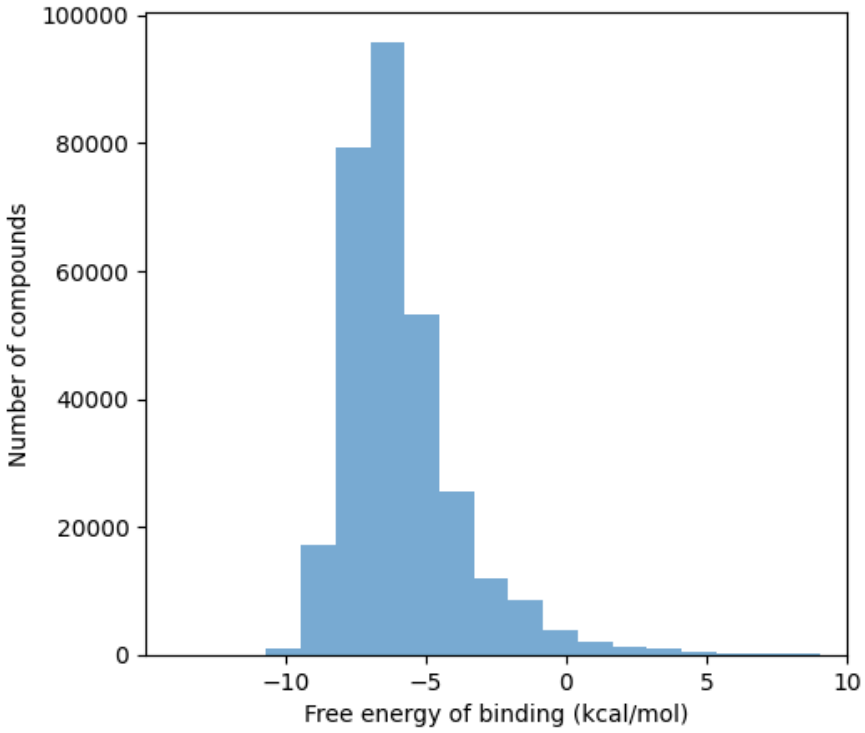

Figure 5. Histogram depicting free energy of binding values obtained for compounds docked in the screening process. Individual results with binding energy above 10 kcal/mol have been omitted.

3.2. Analysis of the most promising results

The best 10 hits of the docking are presented in Table 4. To limit the data, in the next step we focused on compounds with the free energy of binding values below -10 kcal/mol. These are 203 compounds whose binding energy distribution is shown on Figure 6. Those compounds underwent toxicity and synthetic accessibility predictions using the eToxPred.

From the 50 least potentially toxic compounds, the best inhibitor candidates for Molecular dynamics simulation were chosen, taking into account their previous appearance in biological assays documented in the PubChem database[19]. Description of these compounds is shown in Table 5. and their docked positions are depicted on Fig. 7.

PubChem CID	ChEMBL ID	Free energy of binding (kcal/mol)	Known inhibitor activities
168273699	CHEMBL5177768	-11.9	Inhibition of human NLRP3 inflammasome[16]
122588208	CHEMBL4853037	-11.8	Inhibition of human STK4
168273554	CHEMBL5174347	-11.7	Inhibition of human NLRP3 inflammasome[16]
52947883	CHEMBL1241554	-11.6	Antiviral activity against R5 tropic HIV1 BaL[17]
168271618	CHEMBL5175650	-11.5	Inhibition of human NLRP3 inflammasome[16]
135674153	CHEMBL204427	-11.3	Inhibition of human recombinant PDGFRbeta[18]
168279628	CHEMBL5186834	-11.3	Inhibition of human NLRP3 inflammasome[16]
168271146	CHEMBL5172572	-11.2	Inhibition of human NLRP3 inflammasome[16]
122588222	CHEMBL4874985	-11.2	Inhibition of STK4
122588236	CHEMBL4874241	-11.2	Inhibition of wild-type human full length CSNK1G3, Inhibition of STK4

Table 4. Best 10 hits of virtual screening.

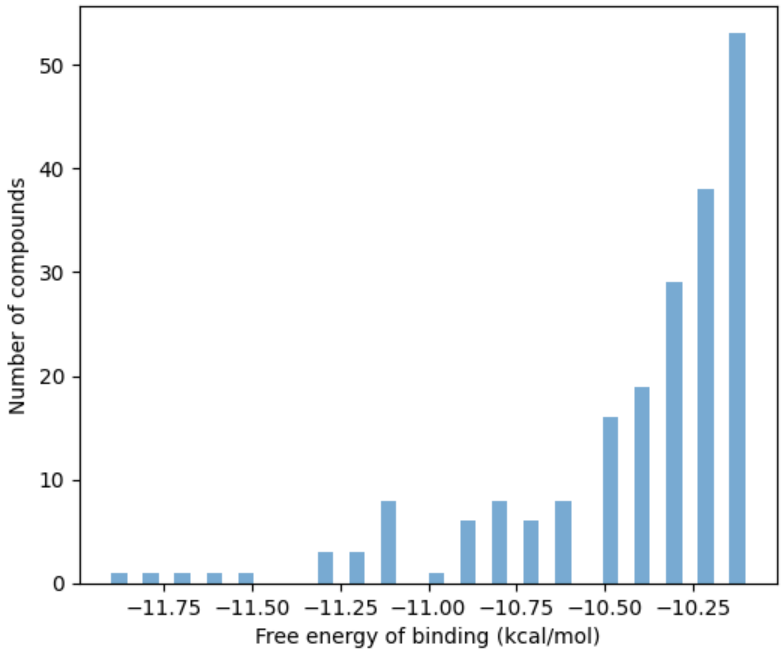

Figure 6. Histogram of the binding free energy values obtained for compounds with values below -10 kcal/mol.

PubChem CID	ChEMBL ID	Free energy of binding (kcal/mol)	Tox-score	SAscore	Known inhibitor activities
52947883	CHEMBL1241554	-11.6	0.54	0.03	Antiviral activity against R5 tropic HIV1 BaL[17]
168290583	CHEMBL5200486	-10.8	0.47	0.09	Inhibition of the bacterial FabI[20]
53322549	CHEMBL1642432	-10.7	0.48	0.28	Potential antifungal and antibacterial activity[21]
1420003	CHEMBL3132949	-10.2	0.44	0.25	Human cancer cell-lines growth inhibition[22]
122187432	CHEMBL3609207	-10.2	0.49	0.16	Antimycobacterial activity against *Mycobacterium tuberculosis* [23]

Table 5. Potential inhibitor candidates: Tox-score entries from 0 meaning less potentially toxic to 1 - probably toxic, SAscore entries from 0 meaning potentially easy to synthetize to 1 - probably difficult to synthesize.

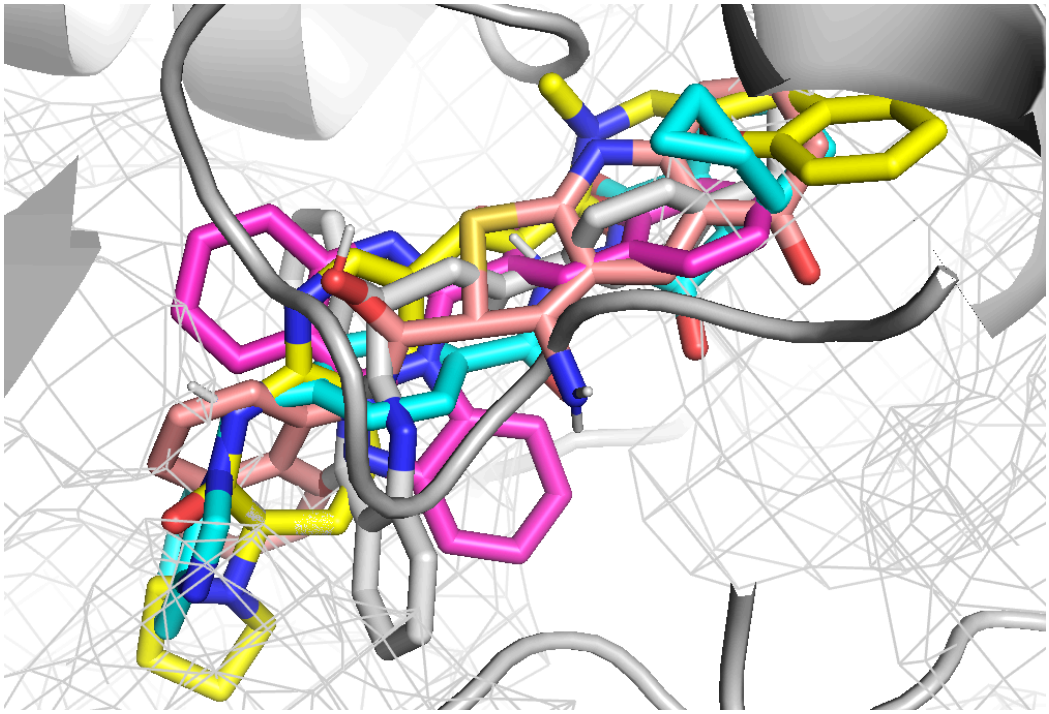

Figure 7. The best inhibitor candidates for Molecular dynamics simulation docked in Nsp13 binding pocket.

4. Discussion

As a docking box for the virtual screening we have chosen an area near the ATP binding site, blocking which potentially would interfere with ATP utilization by Nsp13 helicase necessary for its function. Nevertheless, ATP is an ubiquitous substrate used by numerous enzymes in a human cell, therefore the potential inhibitor interactions with those enzymes should be taken into account.

Interestingly, our best docking hits were mostly inhibitors of different types of human protein kinases, such as Platelet-derived growth factor receptor-β (PDGFR-β), a tyrosine kinase which activates downstream signals resulting in cell proliferation and migration, serine/threonine-protein kinases such as Serine/threonine-protein kinase 4 (Stk4) and Casein kinase I isoform gamma-3 (CSNK1G3) which phosphorylate a large number of proteins and participate in cell signaling, and 5 of the compounds were inhibitors of NLRP3 inflammasome, a multiprotein complex that facilitates activation and release of the proinflammatory cytokines in response to infection or endogenous stimuli.

For example, Vazquez et al. reported that Nsp13 can block interferon activation via distinct mechanisms, among others by binding to TBK1, a kinase mainly known for its role in innate immunity antiviral response.[24] Fung et al. report that Nsp13 interacted with STAT1, a transcription factor involved in interferon signaling, to prevent JAK1 kinase from its activation[25]. It is highly possible that the binding residues we have analyzed are somehow involved in Nsp13-antiviral protein interactions.

Motivated by the concerns about host enzymes-Nsp13 inhibitor interactions, we aimed to choose the best candidate ligands which were not reported to inhibit any human proteins according to PubChem biological assays data. They were reported to have either antiviral, antibacterial/antifungal properties, or inhibit cancer cell growth.

As specified earlier, the main aim of the project was identifying candidates for SARS-CoV-2 Nsp13 inhibitors. It appears that we have obtained preliminary results that undoubtedly need further verification. The next step that should be taken to verify the results is Molecular Dynamics (MD) simulation and flexible docking. We have made several attempts to perform these analyzes using two tools - GROMACS[26] and CHARMM-GUI[27], but unfortunately due to encountered technical obstacles and lack of time we were not able to finish the simulations.

Upon obtaining satisfactory MD results, in the next step it would be worthwhile verifying cell membrane permeability of tested compounds and potential interactions with a human cell enzyme machinery. Moreover, it would be beneficial to further optimize the workflow and even to repeat a process with updated experience and knowledge. In spite of that, we consider the main goal of the project as accomplished.

5. References

1. Raubenolt, B. A., Islam, N. N., Summa, C. M. & Rick, S. W. Molecular dynamics simulations of the flexibility and inhibition of SARS-CoV-2 NSP 13 helicase. *J. Mol. Graph. Model.* **112**, 108122 (2022).
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. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment | Bioinformatics | Oxford Academic. https://academic.oup.com/bioinformatics/article/29/20/2588/277910.
4. PDBe-KB: a community-driven resource for structural and functional annotations | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/48/D1/D344/5580911.
5. Kochnev, Y., Hellemann, E., Cassidy, K. C. & Durrant, J. D. Webina: an open-source library and web app that runs AutoDock Vina entirely in the web browser. *Bioinformatics* **36**, 4513–4515 (2020).

6. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, D1083-1090 (2014).
7. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100-1107 (2012).
8. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26 (2001).
9. RDKit: Open-source cheminformatics.
10. White, M. A., Lin, W. & Cheng, X. Discovery of COVID-19 Inhibitors Targeting the SARS-CoV2 Nsp13 Helicase. *bioRxiv* 2020.08.09.243246 (2020) doi:10.1101/2020.08.09.243246.
11. Tang, S. *et al.* Accelerating AutoDock Vina with GPUs. *Mol. Basel Switz.* **27**, 3041 (2022).
12. 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).
13. NVIDIA, Vingelmann, P. & Fitzek, F. H. P. CUDA, release: 10.2.89. (2020).
14. Pu, L. *et al.* eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacol. Toxicol.* **20**, 2 (2019).
15. Gallicchio, E., Lapelosa, M. & Levy, R. M. The Binding Energy Distribution Analysis Method (BEDAM) for the Estimation of Protein-Ligand Binding Affinities. *J. Chem. Theory Comput.* **6**, 2961–2977 (2010).
16. Harrison, D. *et al.* Discovery and Optimization of Triazolopyrimidinone Derivatives as Selective NLRP3 Inflammasome Inhibitors. *ACS Med. Chem. Lett.* **13**, 1321–1328 (2022).

17. Dong, M. *et al.* Tricyclononene carboxamide derivatives as novel anti-HIV-1 agents. *Eur. J. Med. Chem.* **45**, 4096–4103 (2010).
18. Renhowe, P. A. *et al.* Design, structure-activity relationships and in vivo characterization of 4-amino-3-benzimidazol-2-ylhydroquinolin-2-ones: a novel class of receptor tyrosine kinase inhibitors. *J. Med. Chem.* **52**, 278–292 (2009).
19. PubChem 2023 update | Nucleic Acids Research | Oxford Academic. https://academic.oup.com/nar/article/51/D1/D1373/6777787.
20. Li Petri, G., Di Martino, S. & De Rosa, M. Peptidomimetics: An Overview of Recent Medicinal Chemistry Efforts toward the Discovery of Novel Small Molecule Inhibitors. *J. Med. Chem.* **65**, 7438–7475 (2022).
21. Kuarm, B. S., Reddy, Y. T., Madhav, J. V., Crooks, P. A. & Rajitha, B. 3-[Benzimidazo- and 3-[benzothiadiazoleimidazo-(1,2-c)quinazolin-5-yl]-2H-chromene-2-ones as potent antimicrobial agents. *Bioorg. Med. Chem. Lett.* **21**, 524–527 (2011).
22. The development of thieno[2,3-b]pyridine analogues as anticancer agents applying in silico methods - MedChemComm (RSC Publishing). https://pubs.rsc.org/en/content/articlelanding/2014/md/c3md00320e.
23. Kamal, A. *et al.* Synthesis and biological evaluation of phaitanthrin congeners as anti-mycobacterial agents. *Bioorg. Med. Chem. Lett.* **25**, 3867–3872 (2015).
24. Vazquez, C. *et al.* SARS-CoV-2 viral proteins NSP1 and NSP13 inhibit interferon activation through distinct mechanisms. *PLoS ONE* **16**, e0253089 (2021).
25. Fung, S.-Y. *et al.* SARS-CoV-2 NSP13 helicase suppresses interferon signaling by perturbing JAK1 phosphorylation of STAT1. *Cell Biosci.* **12**, 36 (2022).
26. Bekker, H. *et al.* Gromacs: A parallel computer for molecular dynamics simulations. *Phys.*

Comput. **92**, 252–256 (1993).

27. CHARMM‐GUI: A web‐based graphical user interface for CHARMM - Jo - 2008 - Journal of Computational Chemistry - Wiley Online Library. https://onlinelibrary.wiley.com/doi/10.1002/jcc.20945.