

MeetEU Project - Team Heidelberg - Team 1 –
Identification and Enhancement of novel Sars-CoV-2 NSP13 Helicase
Inhibitors

Linda Blaier, Paul Brunner, Selina Ernst, Valerie Segatz, and Chloé Weiler

February 2024

1 Abstract

Abbreviations

MD	Molecular dynamics
NSP13	Non-structural protein 13
RTC	Replication transcription complex
SAscore	Synthetic accessibility score
ssRNA	Single-stranded RNA
ZBD	Zinc binding domain
SARS-CoV-2	severe acute respiratory syndrome corona virus 2
FDA	food and drug administration
RAS	renin-angiotensin system
COVID-19	coronavirus disease 2019
SARS-CoV-2	SARSI

2 Introduction

Even though the development of vaccines against SARS-CoV-2 was successful during the recent pandemic, the amount of FDA approved drugs for the therapy of COVID-19 is still limited to Paxlovid, Veklury, Olumiant, and Actemra (FDA 2023). Improving the landscape of drugs available for treating COVID-19 would be particularly beneficial for people who are at risk of severe illness, as vaccines may not fully prevent infections. The goal of this year’s Meet-EU project is to develop a pipeline to identify possible inhibitors against the SARS-CoV-2 helicase also known as Non-structural protein 13 (NSP13). There are two main reasons as to why this protein is a promising drug target. For one, it is highly conserved among corona viruses, which means that the virus is unlikely to develop resistances against drugs targeting NSP13 through rapid mutations in the viral genome (Spratt et al. 2021). On the other hand, NSP13 together with other non-structural proteins forms the Replication transcription complex (RTC), which is essential for viral RNA synthesis (Malone et al. 2022). Therefore, inhibiting NSP13 would severely hinder the spread of the virus inside the host. The protein consists of five domains, namely the Zinc binding domain (ZBD), the stalk domain, as well as 1A, 2A and 1B. The latter three make up the catalytic centre of the protein, where RNA and ATP bind (Newman et al. 2021).

Computer-aided structure-based drug discovery can be followed to identify possible inhibitors of the NSP13 helicase. These can then be further investigated in wet-lab settings. This process involves several steps: (1) Identification of possible binding sites, (2) high-throughput screening of ligands for how well they bind the respective pocket, followed by (3) the evaluation of the binding pathways, the kinetics, and thermodynamics (Śledź and Caflisch 2018). Hereby, focussing this screening on well documented or already FDA approved compounds is very attractive, as this drug repurposing potentially shortens the development period and therefore also the development costs (Pushpakom et al. 2019). Single-stranded RNA (ssRNA), SARSI (SARS-CoV-2)

2.1 Identification of Consensus Binding Pocket

In drug discovery, the initial step is to investigate the protein structure in order to analyse potential binding sites. These are cavities on the surface or interior of the protein with suitable properties to bind a ligand. The functionality of a binding pocket is determined by its shape and location, but also by the amino acid residues which define its physicochemical characteristics (Stank et al. 2016). Different experimental and theoretical procedures exist to analyse the druggability of such binding pockets. In this work, we combined three different *in silico* tools, each following a different algorithm. Fpocket (Le Guilloux et al. 2009) utilises a geometry-based algorithm based on Voronoi tessellation and sequential clustering to determine potential binding sites. We also implemented P2Rank (Krivák and Hoksza 2018; Jendele et al. 2019; Jakubec et al. 2022), which is based on a machine-learning algorithm. P2Rank assigns structural, physicochemical, and evolutionary features to points on the solvent-accessible surface of a protein. From this information, the machine-learning model is built and used to predict and rank potential ligand binding sites. Lastly, FTMAP (Brenke et al. 2009) was used to validate the binding pocket found with the previously mentioned approaches. FTMAP uses docking results of sixteen small molecules differing in polarity, shape, and size to identify binding hot spots with a fast Fourier transform correlation. The most favourable docked confirmations are determined through energy minimisation and clustering processes. Finally, the results of all three tools were combined to identify a consensus binding pocket of the NSP13 helicase. The resulting coordinates of the consensus binding pocket were then used for molecular docking simulations.

2.2 Molecular Docking

Molecular Docking programs are used to evaluate binding affinities between a potential drug candidate and the target protein. A key aspect of this task is the prediction of the ligand position, orientation, and conformation. Search-based methods approach this task by continuously modifying the ligand pose, while estimating its quality or likelihood (score) and stochastically trying to infer the global optimum of the scoring function. Among the most widely used tools are AutoDock Vina (Trott and Olson 2010) and Glide (Halgren et al. 2004), which mainly differ in their scoring functions. However, such search-based methods are computationally expensive. Therefore, in order to be able to screen large datasets, search-based methods are generally restricted to a previously defined binding pocket (Corso et al. 2022). Consequently, potential other binding sites of a ligand are not assessed. Machine learning-based blind docking approaches try to address that problem by stochastically predicting binding pocket and ligand pose based on learned characteristics and aligning them. The most promising results are achieved by using Diffdock (Corso et al. 2022), a generative model which applies a reverse diffusion process to the docking paradigm. In this manner, Diffdock iteratively transforms an uninformed noisy distribution over ligand poses defined by the degrees of freedom involved in docking (position, turns around its centre of mass, and twists of torsion angles) into a learned model distribution (Corso et al. 2022). Corso *et al.* thereby describe this process as a progressive refinement of random ligand poses via updates of their translations, rotations and torsion angles.

2.3 Estimation of Toxicity and Synthetic Accessibility

After identifying the lead compounds that exhibit optimal binding affinity within the consensus pocket, an evaluation of the general toxicity and synthetic accessibility of these compounds was performed using the latest version of the machine-learning tool *eToxPred* (Pu et al. 2019). This additional step helps estimate the suitability of the compounds as real-life pharmaceuticals against COVID-19. The Tox-score allows for a general assessment of the predicted risk vs. benefit ratio of the potential NSP13 inhibitors. Moreover, *eToxPred* allows for an insight into the ease of synthesis, indicated by the Synthetic accessibility score (SAscore). This score reflects the ease and efficiency of producing the molecules in large quantities and consequently their feasibility as potential drugs.

2.4 Molecular Dynamics Simulation

As the last step of our pipeline, a Molecular dynamics (MD) simulation is conducted using the best-scoring compound as a ligand in the binding pocket of the NSP13 protein. Using GROMACS (Version 2023.3) (Abraham et al. 2015), this enables us to interpret the stability of the protein-ligand interaction, as well as to identify important residues for the interaction. Using a given force-field, a set of equations describing different forces between the atoms and residues in the protein and ligand, the movement of all atoms in the system can be simulated and analysed. However, this is only possible in a very limited timeframe with a small time step size. As this process is rather resource-heavy, it has to be conducted on a cluster with access to a GPU.

3 Material and Methods

3.1 Toxicity and Synthetic Accessibility Prediction using *eToxPred*

The general toxicity and synthetic accessibility of the given compounds was estimated using the machine-learning tool *eToxPred* (Pu et al. 2019). The SMILES files of the Top100 compounds from AutoDock Vina (Trott and Olson 2010) served as input for the pre-trained model. The toxicity

predictor was pre-trained on the FDA-approved and the KEGG-drug datasets whose compounds were considered non-toxic as well as the TOXNET and the T3DB datasets whose compounds were considered toxic using a deep-belief-network based model. This predictor yields a Tox-score between 0 and 1 and in accordance to the paper, all compounds with a Tox-score below 0.58 were deemed non-toxic. The synthetic accessibility was reflected in a synthetic accessibility score (SAscore) which was obtained by training an extra-trees-based classifier on NuBBE, UNPD, FDA-approved, and DUD-E-active datasets.

4 Results

4.1 Top 100 Compounds exhibit low toxicity and high synthetic accessibility

After identifying the top 100 best-binding ligands through AutoDock Vina, our subsequent analysis focused on evaluating their practical applicability as potential drugs by considering their predicted toxicity and synthetic accessibility. The resulting SAscore and Tox-Score for each compound were visualised in a scatter plot as seen in Figure 1.

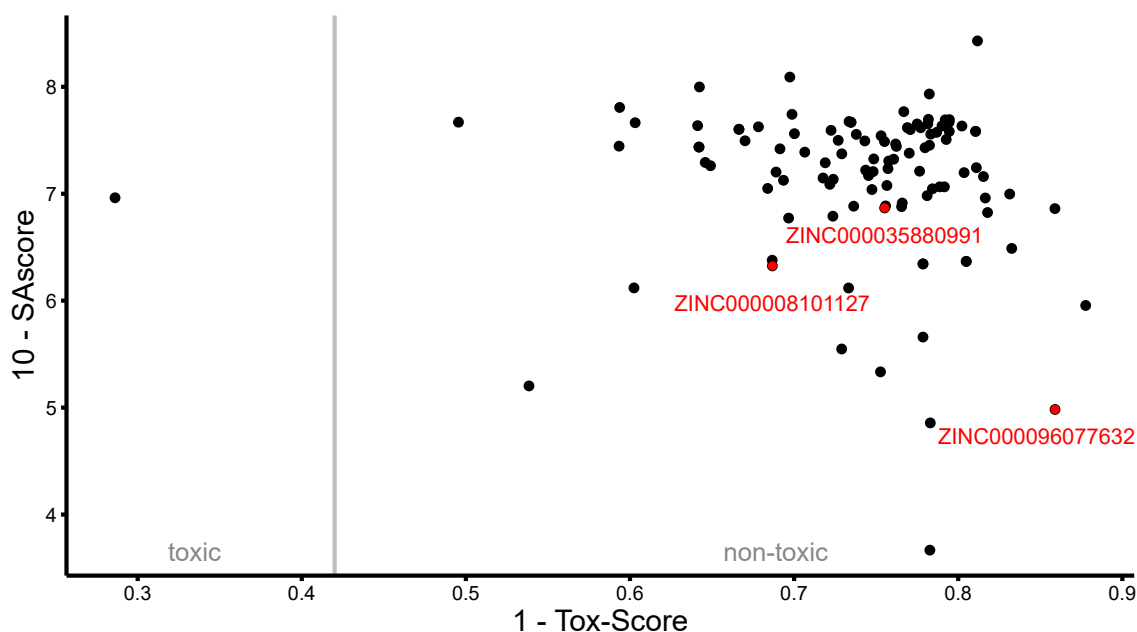


Figure 1: **Scatter plot of predicted toxicity and synthetic accessibility of the 100 best-binding compounds.** The predicted SAscore was plotted against the Tox-Score for the top 100 best scorers from AutoDock Vina. The top 3 scorers from Glide are highlighted in red. The vertical gray line represents the threshold for the toxicity, with the compounds to the right of this line being considered non-toxic.

Of those 100 compounds, 99 presented with a Tox-score below the threshold of toxicity, indicating a low probability of being toxic to humans. The overall median Tox-Score is 0.24, with a mean of 0.26 across all compounds. Across all compounds the median SAscore of 2.69 and a mean of 2.87 which suggests that they are generally easy to synthesise. The top scorer from Glide, ZINC000096077632, was predicted to have a Tox-Score of 0.14 and an SAscore of 5.02.

5 Discussion and Outlook

5.1 toxicity

The Tox-Score predictor of *eToxPred* was trained using food and drug administration (FDA)-approved dataset as non-toxic incidences. Consequently, the low mean Tox-Score of the tested compounds aligns with our expectation, considering that the compounds from the ZINC database are derived from an FDA-approved dataset. The single toxic incidence we detected was from the ECBD database which was comprised of FDA-approved and non-FDA-approved molecules. As highlighted by Pu et al., natural compounds typically exhibit higher SAscore values compared to synthetic compounds due to their inherent complexity Pu et al. 2019. The relatively high SAscore of ZINC000096077632 can be explained by the fact that ZINC000096077632 corresponds to angiotensin-(1-7) which is a naturally occurring compound with a crucial role in the renin-angiotensin system (RAS) Santos 2014. The analysis of natural compound datasets by Pu et al. revealed a bimodal distribution in the SAscore, with peaks around 3 and 5. Furthermore the very low Tox-Score of the top scorer can also be explained by the fact that it is a naturally occurring molecule in the human body.

6 Supplementary Material

References

- Abraham, M. J., T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl (2015). “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. In: *SoftwareX* 1, pp. 19–25. ISSN: 2352-7110. DOI: 10.1016/j.softx.2015.06.001.
- Brenke, R., D. Kozakov, G. Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos, and S. Vajda (2009). “Fragment-based identification of druggable ‘hot spots’ of proteins using Fourier domain correlation techniques”. In: *Bioinformatics* 25.5, pp. 621–7. ISSN: 1367-4803 (Print) 1367-4803. DOI: 10.1093/bioinformatics/btp036.
- Corso, G., H. Stärk, B. Jing, R. Barzilay, and T. Jaakkola (2022). *DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking*. DOI: 10.48550/arXiv.2210.01776.
- FDA (2023). *Know Your Treatment Options for COVID-19*. URL: <https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19>.
- Halgren, T. A., R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks (2004). “Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening”. In: *Journal of medicinal chemistry* 47.7, pp. 1750–1759. ISSN: 0022-2623. DOI: 10.1021/jm030644s.
- Jakubec, D., P. Skoda, R. Krivak, M. Novotny, and D. Hoksza (2022). “PrankWeb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures”. In: *Nucleic Acids Research* 50.W1, W593–W597. ISSN: 0305-1048. DOI: 10.1093/nar/gkac389. URL: <https://doi.org/10.1093/nar/gkac389>.
- Jendele, L., R. Krivak, P. Skoda, M. Novotny, and D. Hoksza (2019). “PrankWeb: a web server for ligand binding site prediction and visualization”. In: *Nucleic Acids Research* 47.W1, W345–W349. ISSN: 0305-1048. DOI: 10.1093/nar/gkz424. URL: <https://doi.org/10.1093/nar/gkz424>.
- Krivák, R. and D. Hoksza (2018). “P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure”. In: *Journal of Cheminformatics* 10.1, p. 39. ISSN: 1758-2946. DOI: 10.1186/s13321-018-0285-8. URL: <https://doi.org/10.1186/s13321-018-0285-8>.
- Le Guilloux, V., P. Schmidtke, and P. Tuffery (2009). “Fpocket: An open source platform for ligand pocket detection”. In: *BMC Bioinformatics* 10.1, p. 168. ISSN: 1471-2105. DOI: 10.1186/1471-2105-10-168. URL: <https://doi.org/10.1186/1471-2105-10-168>.
- Malone, B., N. Urakova, E. J. Snijder, and E. A. Campbell (2022). “Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design”. In: *Nature Reviews Molecular Cell Biology* 23.1, pp. 21–39.
- Newman, J. A., A. Douangamath, S. Yadzani, Y. Yosaatmadja, A. Aimon, J. Brando-Neto, L. Dunnett, T. Gorrie-stone, R. Skyner, D. Fearon, M. Schapira, F. v. Delft, and O. Gileadi (2021). “Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase”. In: *Nature Communications* 12.1, p. 4848. DOI: 10.1038/s41467-021-25166-6.
- Pu, L., M. Naderi, T. Liu, H.-C. Wu, S. Mukhopadhyay, and M. Brylinski (2019). “e toxpred: A machine learning-based approach to estimate the toxicity of drug candidates”. In: *BMC Pharmacology and Toxicology* 20, pp. 1–15.
- Pushpakom, S., F. Iorio, P. A. Eyers, K. J. Escott, S. Hopper, A. Wells, A. Doig, T. Guilleams, J. Latimer, C. McNamee, A. Norris, P. Sanseau, D. Cavalla, and M. Pirmohamed (2019). “Drug repurposing: progress, challenges and recommendations”. In: *Nature Reviews Drug Discovery* 18.1, pp. 41–58. ISSN: 1474-1784. DOI: 10.1038/nrd.2018.168. URL: <https://doi.org/10.1038/nrd.2018.168>.
- Santos, R. A. (2014). “Angiotensin-(1–7)”. In: *Hypertension* 63.6, pp. 1138–1147.

- Śledź, P. and A. Caflisch (2018). “Protein structure-based drug design: from docking to molecular dynamics”. In: *Current Opinion in Structural Biology* 48, pp. 93–102. ISSN: 0959-440X. DOI: <https://doi.org/10.1016/j.sbi.2017.10.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0959440X17301100>.
- Spratt, A. N., F. Gallazzi, T. P. Quinn, C. L. Lorson, A. Sönnnerborg, and K. Singh (2021). “Coronavirus helicases: Attractive and unique targets of antiviral drug-development and therapeutic patents”. In: *Expert opinion on therapeutic patents* 31.4, pp. 339–350.
- Stank, A., D. B. Kokh, J. C. Fuller, and R. C. Wade (2016). “Protein Binding Pocket Dynamics”. In: *Accounts of Chemical Research* 49.5. doi: 10.1021/acs.accounts.5b00516, pp. 809–815. ISSN: 0001-4842. DOI: 10.1021/acs.accounts.5b00516. URL: <https://doi.org/10.1021/acs.accounts.5b00516>.
- Trott, O. and A. J. Olson (2010). “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading”. In: *Journal of computational chemistry* 31.2, pp. 455–461. DOI: 10.1002/jcc.21334.