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

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 and Veklury, Olumiant and Actemra FDA 2023. One possibility to accelerate the development of new therapies for Covid19 is to screen already approved drugs for effects against the viral reproduction. In this years MeetEU project, we investigated the NSP13 helicase of Sars-CoV-2 and tried to find compounds that could be repurposed for this therapy, as well as novel compounds that could lead to an effective treatment of Covid19. Using our *in-silico* pipeline enables us to evaluate possible drug candidates, suggest novel structures based on already approved drugs and investigate their toxicity, while being cheaper and less labor intensive than projects limited to wet-lab work. HIER KOMMT EINE ABBILDUNG HIN

2 Introduction

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 pyhsicochemical 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 tesselation and sequential clustering to determine potential binding sites. We also implemented P2Rank Kriv´ak and Hoksza 2018; Jendele et al. 2019; Jakubec et al. 2022 which is based on a machine-learning algorithm. P2Rank assigns structural, physico-chemical, 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 binidng 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 to 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 function. 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 thereby achieved by using Diffdock, a generative model which applies a reverse diffusion process to the docking paradigm. Thereby, 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 Lead Drug Enhancement

In order to enhance the binding affinity of our drug candidates and thus their performance, we used AutoGrow4 (Version 4.0.3) Spiegel and Durrant 2020 to generate novel compounds. Starting with the best binding compounds of our initial docking simulation with AutoDock Vina as generation zero, multiple new structures are generated by combining sub-structures of the first generation or by passing them through a set of possible chemical reactions after converting them into their respective SMILES codes. All of the generated compounds are ranked by their binding affinity. After passing several filters, the best-performing compounds are used as the seed for the next generation. Using this algorithm, compounds are found, which show higher binding affinities than the first generation. As AutoGrow4 labels all new structures by the path by which they were obtained, we can also evaluate the synthesizability.

2.4 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 *e*ToxPred (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, *e*ToxPred allows for an insight into the ease of synthesis, reflected by the synthetic accessibility score (SAscore), of the compounds in a laboratory setting.

2.5 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 *e*ToxPred

Following lead optimisation, the toxicity and synthetic accessibility of the given compounds was estimated using the machine-learning tool *e*ToxPred (Pu et al. 2019) **on molecular fingerprints**. The

toxicity preditor was pre-trained on the FDA-approuved 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 are 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-approuved, and DUD-E-active datasets. As input for the pre-trained model served the SMILES files of the Top100 compounds from AutoDock Vina **CITATION**.

4 Results

5 Discussion and Outlook

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.

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.

Spiegel, J. O. and J. D. Durrant (2020). "AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization". In: *Journal of Cheminformatics* 12.1, p. 25. ISSN: 1758-2946. DOI: 10.1186/s13321-020-00429-4.

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.