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 ([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.,]). 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.,]), 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.,]). 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 *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, 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 Datasets from ZINC20 and ECBL

A total of 1616 fda approved drugs were downloaded in .sdf format from the ZINC database ([Irwin et al., 2020]). Additionally, 5016 files were retrieved, downloading the pilot library from the ECBL database.

3.2 Receptor and Ligand Preparation

Ligands were prepared using openbable in order to convert implicit hydrogens into explicit hydrogens, generate necessary 3D structures of the ligands, as well as to split mulitmolecule files into single ones. ADFR suite was further used in order to convert all files into the .pdbqt format, which is required by Autodock Vina.

4 Results

5 Discussion and Outlook

6 Supplementary Material

References

[Abraham et al., 2015] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25.

[Brenke et al., 2009] Brenke, R., Kozakov, D., Chuang, G. Y., Beglov, D., Hall, D., Landon, M. R., Mattos, C., and Vajda, S. (2009). Fragment-based identification of druggable 'hot spots' of proteins using fourier domain correlation techniques. *Bioinformatics*, 25(5):621–7.

[Corso et al.,] Corso, G., Stärk, H., Jing, B., Barzilay, R., and Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking.

[FDA, 2023] FDA (2023). Know your treatment options for covid-19.

[Halgren et al., 2004] Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of medicinal chemistry*, 47(7):1750–1759.

[Irwin et al., 2020] Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., and Sayle, R. A. (2020). Zinc20-a free ultralarge-scale chemical database for ligand discovery. *Journal of chemical information and modeling*, 60(12):6065–6073.

[Jakubec et al., 2022] Jakubec, D., Skoda, P., Krivak, R., Novotny, M., and Hoksza, D. (2022). Prankweb 3: accelerated ligand-binding site predictions for experimental and modelled protein structures. *Nucleic Acids Research*, 50(W1):W593–W597.

[Jendele et al., 2019] Jendele, L., Krivak, R., Skoda, P., Novotny, M., and Hoksza, D. (2019). Prankweb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Research*, 47(W1):W345–W349.

[Krivák and Hoksza, 2018] Krivák, R. and Hoksza, D. (2018). P2rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39.

[Le Guilloux et al., 2009] Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1):168.

[Pu et al., 2019] Pu, L., Naderi, M., Liu, T., Wu, H.-C., Mukhopadhyay, S., and Brylinski, M. (2019). e toxpred: A machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacology and Toxicology*, 20:1–15.

[Stank et al., 2016] Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016). Protein binding pocket dynamics. *Accounts of Chemical Research*, 49(5):809–815. doi: 10.1021/acs.accounts.5b00516.

[Trott and Olson, 2010] Trott, O. and Olson, A. J. (2010). 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.