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 Covid-19 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 Covid-19. 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). There are both different experimental and theoretical procedures existent 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 seqeuntial 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-learnign algorithm. It 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 other two approaches. This tool uses docking results of sixteen small molecules differing ins polarity, shape and size to identify binidng hot spots with a fast Fourier transform correlation. The most favorable docked confirmations are determined by energy minimization and clustering. Finally, the results of the three approaches 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 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.

2.3 Molecular Dynamics Simulation

As the last step of our pipeline, a MD simulation is conducted using the best scoring compounds 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 Protein preparation

In this project we used the crystal structures of the NSP13 helicase of SARS-CoV-2. They were obtained by **Newman et al.** in a crystallographic fragment screening. Thus, three protein structures were downloaded from the protein data bank RCSB PDB (PDB codes: 6ZSL, 5RME, 5RM2). Write more on resuolution and why 6zsl was chosen for further work...
Although the helicase is a homodimer, it was found that only the monomer is the catalytically active form. Therefore, we concentrated our drug discovery only on the monomer of the helicase, namely on Monomer A.

From the given PDB structures

3.2 Consensus binding site detection

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.

FDA (2023). *Know Your Treatment Options for COVID-19*. URL: https://www.fda.gov/consumers/consumer-updates/know-your-treatment-options-covid-19.

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.

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.