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.

Abbreviations

COVID-19   coronavirus disease 2019

FDA   food and drug administration

MD   molecular dynamics

NSP13   non-structural protein 13

RTC   replication transcription complex

SARS-CoV-2   severe acute respiratory syndrome coronavirus type 2

SAscore   synthetic accessibility score

ssRNA   single-stranded RNA

ZBD   zinc-binding domain

2   Introduction

In the wake of the successful development of vaccines against severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) during the recent pandemic, the amount of FDA-approved drugs for the therapy of coronavirus disease 2019 (COVID-19) remain limited to Paxlovid, Veklury, Olumiant, and Actemra (FDA 2023). Expanding the repertoire of drugs available for treating COVID-19 is particularly crucial for high-risk individuals, as vaccines may not fully prevent infections. The goal of this year's Meet-EU project is to develop a pipeline for identifying potential inhibtiors targeting the SARS-CoV-2 helicase also known as non-structural protein 13 (NSP13). This protein is considered a promising drug target for two reasons. Firstly, NSP13 is part of the replication transcription complex (RTC) which is essential for viral RNA synthesis (Malone et al. 2022). Secondly, its high conservation across coronaviruses means that there is a reduced likelihood of the virus developing resistances to drugs targeting NSP13 through mutations of the viral genome (Spratt et al. 2021). Consequently, effective NSP13 inhibition would significantly impede viral replication and therefore its spread within the host. NSP13 consists of five domains, namely the zinc-binding domain (ZBD) with the accessory 2B domain, the α-helical stalk domain, as well as the RecA-like 1A and 2A domains (Marecki et al. 2021). The gap between the RecA-like domains and the 2B domain encompass the site for ATP and RNA binding (Newman, Douangamath, Yadzani, Yosaatmadja, Aimon, Brando-Neto, et al. 2021).

Computer-aided, structure-based drug discovery is employed for identification of known drugs for their potential as inhibitors of the NSP13 helicase which can then be further investigated in experimental laboratory settings. This process involves several steps: (1) Identification of potential drug binding sites, (2) high-throughput screening of ligands to assess their binding efficacy within the designated pocket, followed by (3) the evaluation of the binding pathways, kinetics, and thermodynamics (Śledź and Caflisch 2018). Focussing this screening on well-documented or already FDA-approved compounds is very attractive, as drug repurposing has the potential to shorten the development period and therefore also the development costs (Pushpakom et al. 2019).

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 (Version 3.0, Le Guilloux et al. (2009)) utilises a geometry-based algorithm based on Voronoi tesselation and sequential clustering to determine potential binding sites. We also used PRankweb which is the web interface based on the P2Rank standalone method which is based on a machine-learning algorithm (Krivák and Hoksza 2018; Jendele et al. 2019; Jakubec et al. 2022). 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, the latest version of 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. (2022) 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 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 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

In this project we analysed crystal structures of the SARS-CoV-2 NSP13 helicase which were obtained by (Newman, Douangamath, Yadzani, Yosaatmadja, Aimon, Brandão-Neto, et al. 2021) in a crystallographic fragment screening. Thus, the three protein structures (PDB codes: 6ZSL, 5RME, 5RM2) were downloaded from the protein data bank RCSB PDB. All three structures were used to obtain a consensus binding pocket. Since, 6ZSL is the crystal structure with the highest resolution of 1.94 Å and both other structures show the NSP13 helicase in complex with a different fragment, we selected 6ZSL for further analysis in our drug discovery pipeline. 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 chain A.

Protein Data Bank files often contain problems that first need to be resolved so they can be simulated. Therefore, we prepared the PDB protein structure using the PDBFixer (Version 1.9) application, which among other things adds missing hydrogen and heavy atoms, builds missing loops and replaces non-canonical with canonical amino acids (Eastman et al. 2017). Here, we added hydrogen atoms appropriate for the physiological pH of 7.4. Since the catalytically active form of NSP13 helicase is the monomer (Berta et al. 2021), the additional second copy (Chain B) contained in the crystallographic structure, was removed as well as the ligands contained in the protein structures 5RME and 5RM2. Additionally, for pocket detection the phosphates and zinc ions were removed using PyMol (Version 2.5.7). 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.

3.3 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 *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.

3.4 Molecular Dynamics Simulation

In order to validate the binding of the discovered compounds, we used GROMACS (Version 2023.3, **packageGROMACS**) to simulate the drugs inside of the ATP binding site of NSP13. To do so, the Protein-Ligand Complex tutorial by Lemkul was followed (Lemkul 2018). The a99SB-disp forcefield was used, as it was shown to recreate protein structures in different environments very accurately (Robustelli et al. 2018). The 6ZSL PDB file was first turned into the fitting GROMACS format

(*.gro*). Using the *.gro* file, the general simulation parameters are set up. This includes setting the size and shape of the simulation box, which was set to be a dodecahedron in this case.

As the ligands feature bonds and atoms not commonly seen in proteins, it is required to create a fitting force-field for them. For this, acpype (**acpype**) was used, which was developed to be compatible with AMBER force-fields. The output files from Maestro Glide were converted to PDB and then piped into acpype. The software uses the given PDB file and information about the charges on the compound to create multiple files for different simulation suites, that enable us to add the compound to the simulation environment with the right descriptions of its atoms. The output of acpype was then combined with the GROMACS topology and *.gro* files manually, according to Lemkul.

Following the tutorial, the simulation space was created, filled with water (a99SBdisp_water used as its force-field) and ions were added to create a net zero charge system. As it is common the charges were equalized using sodium and chloride ions. Energy-minimization was conducted and *NVT* and *NPT* equilibration was used to finish the simulation setup. The V-Rescale thermostat was used for thermal coupling in the *NVT* equilibration and for the *NPT* equilibration, a Parrinello-Rahman pressure coupling was used. This ensures a somewhat stable starting point for the simulation. For these steps, the protein and ligand were energetically coupled, as they are in close proximity to each other. The rest of the simulation space, namely ions and water, were also coupled. The production run was set to simulate 100 ns, using a Leap-Frog Integrator. The *.mdp* file containing the simulation parameters used for the final production run can be seen in the appendix. One of these runs took approximately 36 h, which was made possible by parallelization and usage of a NVIDIA V100 GPU. For the analysis of the simulation, GROMACS internal tools were used to calculate the RMSD and RMSF of the protein. These scores give us insight into the movement and conformational changes during the simulation regarding protein and ligand. Furthermore, the simulation frames were extracted and rendered into a video using VMD (**VMD**). For MM-PBSA calculations we used gmx_MMPBSA (**MMPBSA1**). This package enables direct calculation of binding affinities using the GROMACS output files.

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). "GRO-MACS: 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.

Berta, D., M. Badaoui, S. A. Martino, P. J. Buigues, A. V. Pisliakov, N. Elghobashi-Meinhardt, G. Wells, S. A. Harris, E. Frezza, and E. Rosta (2021). "Modelling the active SARS-CoV-2 helicase complex as a basis for structure-based inhibitor design". In: *Chemical Science* 12.40, pp. 13492–13505.

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.

Eastman, P., J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande (2017). "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics". In: *PLOS Computational Biology* 13.7, e1005659. DOI: 10.1371/journal.pcbi.1005659. URL: https://doi.org/10.1371/journal.pcbi.1005659.

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.

Irwin, J. J., K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, and R. A. Sayle (2020). "ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery". In: *Journal of chemical information and modeling* 60.12, pp. 6065–6073. DOI: 10.1021/acs.jcim.0c00675.

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.

Lemkul, J. A. (Oct. 2018). "From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]". In: *Living Journal of Computational Molecular Science* 1.1, p. 5068. DOI: 10.33011/livecoms.1.1.5068. URL: https://livecomsjournal.org/index.php/livecoms/article/view/v1i1e5068.

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.

Marecki, J. C., B. Belachew, J. Gao, and K. D. Raney (2021). "Chapter Ten - RNA helicases required for viral propagation in humans". In: *Viral Replication Enzymes and their Inhibitors Part B*. Ed. by C. E. Cameron, J. J. Arnold, and L. S. Kaguni. Vol. 50. The Enzymes. Academic Press, pp. 335–367. DOI: `https://doi.org/10.1016/bs.enz.2021.09.005`. URL: `https://www.sciencedirect.com/science/article/pii/S1874604721000226`.

Newman, J. A., A. Douangamath, S. Yadzani, Y. Yosaatmadja, A. Aimon, J. Brandão-Neto, L. Dunnett, T. Gorrie-stone, R. Skyner, D. Fearon, M. Schapira, F. von 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. ISSN: 2041-1723. DOI: `10.1038/s41467-021-25166-6`. URL: `https://doi.org/10.1038/s41467-021-25166-6`.

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. Guilliams, 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`.

Robustelli, P., S. Piana, and D. E. Shaw (2018). "Developing a molecular dynamics force field for both folded and disordered protein states". In: *Proceedings of the National Academy of Sciences* 115.21, E4758–E4766. ISSN: 0027-8424. DOI: `10.1073/pnas.1800690115`.

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. Sönnerborg, 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`.