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˙COVID**). 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

MD   Molecular dynamics

NSP13   Non-structural protein 13

RTC   Replication transcription complex

SAscore   Synthetic accessibility score

ssRNA   Single-stranded RNA

ZBD   Zinc binding domain

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˙COVID**). 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 inhibtiors against the SARS-CoV-2 helicase also known as **NSP13!** (**NSP13!**). There are two main reasons, 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˙2021**). On the other hand, **NSP13!** together with other non-structural proteins forms the **RTC!** (**RTC!**), which is essential for viral RNA synthesis (**Malone˙2022**). Therefore, inhibiting **NSP13!** would severely hinder the spread of the virus inside the host. The protein consists of five domains, namely the **ZBD!** (**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 (**NSP13˙basics**).

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-trhoughput screening of ligands for how well they bind the respective pocket, followed by (3) the evaluation of the binding pathways, the kinetics, and thermodynamics (**Sledz˙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˙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˙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, **package˙Fpocket**) 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 (**package˙P2Rank**; **package˙PrankWeb**; **package˙PrankWeb3**). 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 (**package˙FTMAP**) 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.2010**) and Glide (**Halgren.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.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.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.2022**). **Corso.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 (**pu2019toxpred**). 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! (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

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 **Lemkul2018** was followed (**Lemkul2018**). The a99SB-disp forcefield was used, as it was shown to recreate protein structures in different environments very accurately (**Forcefield**). 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 per hand, according to **Lemkul2018**.

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. The *.mdp* file containing the simulation parameters used for the final production run can be seen in the appendix. One of these runs approximately took 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 g_mmpbsa (**MMPBSA1**; **MMPBSA2**). This package enables direct calculation of binding affinities using the GROMACS output files.

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.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