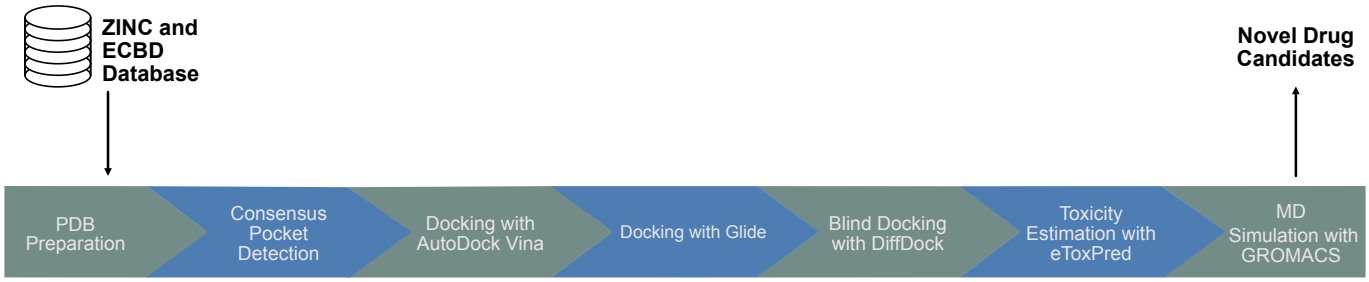
MeetEU Project - Team Heidelberg - Team 1 – Identification and Enhancement of novel Sars-CoV-2 NSP13 Helicase Inhibitors

Linda Blaier, Paul Brunner, Selina Ernst, Valerie Segatz, and Chloé Weiler

February 2024

1 Abstract

The development of innovative therapeutics is essential for preventing another COVID-19 pandemic. Consequently, researchers aim to identify anti-SARS-CoV-2 agents as substitutes for vaccines or immune therapeutics that are able to maintain their efficacy despite the virus's high mutation rate. Targeting the NSP13 helicase of SARS-CoV-2 is a promising strategy due to the protein's high sequence conservation and indispensability for viral proliferation as part of the replication and transcription complex (Marecki et al. 2021; Malone et al. 2022). Therefore, this project focused on the screening of potential drugs that inhibit the NSP13 helicase by trying to find novel compounds as well as FDA-approved compounds that could be repurposed for an effective treatment of COVID-19. In summary, our proposed *in silico* pipeline, depicted in the graphical abstract, identified the ATP-binding pocket of NSP13 as the consensus binding site. Notably, we identified angiotensin 1-7 as our top scoring ligand, which has a low predicted toxicity and binds stably inside the binding pocket. This finding suggests the potential of the identified protein as an inhibitor of NSP13, warranting further *in vitro* validation.



Proposed workflow for the discovery of NSP13 helicase inhibitors.

Abbreviations

ADP	adenosine diphosphate
ATP	adenosine triphosphate
COVID-19	coronavirus disease 2019
DScore	druggability score
FDA	food and drug administration
MD	molecular dynamics
MM-PBSA	molecular mechanics energies combined with the Poisson-Boltzmann and surface area continuum solvation
NSP13	non-structural protein 13
RAS	renin-angiotensin system
RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Data Bank
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
RTC	replication and transcription complex
SARS-CoV-2	severe acute respiratory syndrome coronavirus type 2
SAscore	synthetic accessibility score
ssRNA	single-stranded RNA
Vina	AutoDock Vina
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 inhibitors targeting the SARS-CoV-2 helicase 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 and transcription complex (RTC) which is essential for viral RNA synthesis (Malone et al. 2022). Secondly, its high sequential and structural 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 et al. 2021). In the RTC the NSP13 helicase is present as a homodimer. Nevertheless, only one of the copies complexes with a single-stranded RNA single-stranded RNA (ssRNA). Therefore, previous publications proposed that the monomer is catalytically active form of the NSP13 helicase (Berta 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: (i) identification of potential drug binding sites, (ii) high-throughput screening of ligands to assess their binding efficacy within the designated pocket, and finally (iii) 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 of new therapeutics (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 binding sites are cavities on the surface or the 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 physico-chemical characteristics (Stank et al. 2016). Different experimental and theoretical procedures exist to analyse the druggability of such binding pockets. Overall, merging distinct approaches enables a more precise prediction of the resulting consensus binding pocket which can then be further investigated using molecular docking (Ricci et al. 2022).

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 iteratively modifying the ligand pose. A scoring function summarises the likelihood and quality of each pose by predicting the binding free energy and consequently the pose with the best score is chosen. 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 the binding pocket and

ligand pose based on learned characteristics and aligning them to each other. 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. 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

In addition to determining the activity of novel drug candidates on the therapeutic target, predicition of toxic effects is an indispensible step in drug design to be able to assess the predicted risk vs. benefit ratio of the potential drug (Roncaglioni et al. 2013). As conventional *in vivo* animal tests are time-consuming, expensive, and ethically controversial, researchers nowadays tend to favour *in silico* methods as they are significantly cheaper and faster than wet expirements and they allow for simultaneous evaluation of large numbers of potential drug candidates (Raies and Bajic 2016; Roncaglioni et al. 2013). Therefore, *in silico* toxicity tests are routinely integrated into the early stages of drug discovery in an attempt to minimise late-stage failures in drug design (Dearden 2003). Moreover, novel drugs must not only ensure the safety of patients but also have the capability for large-scale synthesis in order to one day be commercially viable. For that reason, determination of the synthetic accessibility, that is the ease of synthesis of a chemical compound, is essential for estimating the feasibility of an active compound as a pharmaceutical (Boda et al. 2007). Therefore, in our pipeline we followed up the identification of the lead compounds that exhibit optimal binding affinity within the consensus pocket with an evaluation of the general toxicity and synthetic accessibility of these compounds to help estimate the suitability of the compounds as real-life pharmaceuticals against SARS-CoV-2.

2.4 Molecular Dynamics Simulation

As the last step in our pipeline, a molecular dynamics (MD) simulation was be conducted using the best-scoring compounds as a ligand in the binding pocket of the protein. Simulating the movement of molecules in a system based on the attractive and repulsive forces between atoms helps with understanding the molecular interplay between the found ligands and the binding pocket. Whether or not the ligand stays inside the binding pocket throughout the whole simulation gives an estimation of how strongly it binds to the protein (Durrant and McCammon 2011). As this is the final analysis step, the result of this simulation estimates how the ligand would perform in clinical applications. Additionally, the frames generated by the software can be used to calculate different metrics regarding the binding strength, like the root-mean-square fluctuation (RMSF), root-mean-square deviation (RMSD) and the binding affinities estimated by molecular mechanics energies combined with the Poisson-Boltzmann and surface area continuum solvation (MM-PBSA). The RMSD measures the average displacement of the atoms throughout the simulation compared to the first frame of the simulation and estimates how much the protein moves and changes conformations over time. The RMSF on the other hand calculates the movement of a certain atom or group of atoms over time compared to the average position of the atoms and groups (Martinez 2015). The MM-PBSA is able to give an estimation of the binding affinity of the simulated protein-ligand pair (Genheden and Ryde 2015).

3 Material and Methods

3.1 Datasets from ZINC20 and ECBD

A total of 1616 food and drug administration (FDA)-approved drugs were downloaded in *.sdf* format from the ZINC database (Irwin et al. 2020). Additionally, 5016 files of the **ECBD!** (**ECBD!**) pilot library were downloaded.

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 et al. 2021) in a crystallographic fragment screening. Thus, the three protein structures (PDB codes: 6ZSL, 5RME, 5RM2) were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB). All three structures were used to determine a consensus binding pocket. As 6ZSL is the crystal structure with the highest resolution of 1.94 Å and the other two structures show the NSP13 helicase complexed 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 (Berta et al. 2021). Therefore, we concentrated our drug discovery only on the monomer of the helicase, namely on chain A.
Protein Data Bank (PDB) files often contain problems that first need to be resolved, so they can be used for further simulations. 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. The B chain of 6ZSL, identical to the A chain, contained in the crystallographic structure was removed as well as all phosphates and the ligands contained in the protein structures of 5RME and 5RM2. Additionally, for molecular docking using *AutoDock Vina* and for the MD simulation the zinc ions were removed. For consensus binding site detection 6ZSL was used as a reference structure to align 5RME and 5RM2 in *PyMol* (Version 2.5.7, Schrödinger, LLC (2023)). This allowed better visualization and superimposition of the results.
Ligands were prepared using *OpenBabel* (Version 3.1.1, O'Boyle et al. (2011)) in order to convert implicit hydrogens into explicit hydrogens, to generate the necessary 3D structures of the ligands, and to split multi-molecule files into single ones. *ADFR suite* (Version 1.0, Ravindranath et al. (2015)) was further used in order to convert all files into the *.pdbqt* format, which is required by *Autodock Vina* (Trott and Olson 2010).

3.3 Consensus binding site detection

To identify a potential ligand binding site of the NSP13 helicase, three different tools based on different methods were jointly used on Chain A of the crystal structures 6ZSL, 5RM2, and 5RME: (i) *Fpocket* (Version 3.0, Le Guilloux et al. (2009)), (ii) *PrankWeb 3* (accessed on 27.12.2023, Krivák and Hoksza (2018), Jendele et al. (2019), and Jakubec et al. (2022)), and (iii) *FTMap* (accessed on 7.12.2023, Brenke et al. (2009)). *Fpocket* utilizes a geometry-based algorithm based on Voronoi tessellation and sequential clustering to determine potential binding sites (Le Guilloux et al. 2009). Then, for each pocket different properties from the atoms of the pocket are calculated and a druggability score (DScore) is assigned based on which the pockets are ranked (Le Guilloux et al. 2009). The DScore is a weighted sum of the normalized mean local hydrophobic density, a polarity score and a hydrophobicity score to lower the score of pockets where non druglike molecules could bind (Ricci et al. 2022). In this project only pockets with a DScore above 0.2 were considered. We also implemented *PrankWeb 3* which is based on a machine-learning algorithm *P2Rank*. 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 based on a calculated cumulative ligandability score. It also calculates a probability for each pocket which is based on the confidence of the prediction. The results of both tools were visualized in *PyMol*. Then, overlapping hot spots were merged based on the overlapping amino acid sequence and by means of visual inspection. The third tool we used is *FTMap*. It 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 by energy minimization and clustering. Since there is no relatable druggability scoring function as found in both the other tools, we used *FTMap* for validation to see if indeed clusters of the small molecules formed inside the potential consensus binding pocket. Lastly, the centre coordinates of the consensus binding pocket needed for molecular docking were calculated using the public server at usegalaxy.org of the Galaxy web platform

(Afgan et al. 2016). The docking box size was set to have an edge length of 30 Å to ensure that all amino acids of the pocket that could potentially interact with a ligand are considered and that larger ligands are included in the molecular docking simulation.

3.4 Molecular Docking

The molecular docking was performed twice. For the first step *AutoDock Vina* (Version 1.1.2, henceforth abbreviated as Vina, Trott and Olson (2010)) was utilized. The consensus pocket was introduced as the grid box with lengths of 30 Å. The exhaustiveness was set to 30 and the maximum number of binding modes to 9. Taking advantage of multithreading, AutoDock Vina (Vina) used the 28 CPUs accessible on the multi-core server (Che et al. 2023). A filter was applied on the set of ligands assuring only 3D structures smaller than the specified grid box were screened against the receptor The filter was implemented in Python (Version 3.11.6, Van Rossum and Drake (2009)) and executed together with the Vina command in Bash script. The resulting nine different conformations for each ligand were ranked by their binding affinity and only the best value was considered in further steps. A number of ligands were later found to have multiple docking results due to an overlap between the two datasets and in accordance with previous steps only the best score was kept. The remaining 4863 ligands were ranked by their affinity score and the top one hundred were selected for next steps.
A second molecular docking was performed with those top scorers from the screening as well as adenosine diphosphate (ADP) and adenosine triphosphate (ATP). The docking software *Glide* provided by *Schrödinger Inc.* (Friesner et al. 2004) was accessed through *Maestro* (Version 2022.3, Schrödinger Release 2022-3). The included tools *Protein Preparation Wizard* and *LigPrep* (Madhavi Sastry et al. 2013) were utilized to prepare the monomer helicase and ligands for the docking process with the OPLS4 force field (Lu et al. 2021). The zinc ions from the original protein structure were retained. The pH value was set to 7.0. Depending on the initial structure the ligand preparation generates a varying amount of conformations. In the analysis of the results only the best-performing conformation was included. The *Receptor Grid Generation* tool was used to generate the receptor grid with the same binding pocket as in Vina. The docking with Glide was performed at standard precision mode and with flexibility of the ligands enabled (Halgren et al. 2004). The criteria for the selection of the best-performing ligands was chosen to be the docking score. The interactions between the top scoring ligands and the receptor were documented.

3.5 Estimation of Toxicity and Synthetic Accessibility

The general toxicity and synthetic accessibility of the given compounds was estimated using the machine-learning tool *eToxPred* (Pu et al. 2019). The SMILES files of the Top100 compounds from *AutoDock Vina* (Trott and Olson 2010) served as input for the pre-trained model. The toxicity predictor was pre-trained on the FDA-approved 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 were 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-approved, and DUD-E-active datasets.

3.6 Validation of the binding-site for the Top 100 scorer

The likelihood of our Top100 scorer binding within our consensus pocket was validated by blind docking via *Diffdock* (Corso et al. 2022). Thereby, all required sdf files were transformed into SMILES using *OpenBabel* (Version 3.1.1, O'Boyle et al. (2011)). As receptor, the prepared 6ZSL PDB files were utilized. *Diffdock* was performed using the pre-trained scoring model[1] to infer ligand conformations and ranking based on the included confidence model[2]. The proposed default settings were used for the inference. These contained 20

[1] https://github.com/gcorso/DiffDock/tree/bc6b5151457ea5304ee69779d92de0fded599a2c/workdir/paper_score_model
[2] https://github.com/gcorso/DiffDock/tree/bc6b5151457ea5304ee69779d92de0fded599a2c/workdir/paper_confidence_model

inference steps, the generation of 40 samples per ligand, a batch-size of 10 and 18 actual denoising steps, whereby no noise was used for the final step of the reverse diffusion process. For the analysis of the *Diffdock* output, the maximum and minimum coordinates for all samples generated per ligand were extracted using *PyMOL* (Version 2.5.5, Schrödinger, LLC (2015)) and aligned with the coordinates of our consensus pocket in Python (Version 3.10, Van Rossum and Drake (2009)). Based on that, the mean number of ligands within our binding site was calculated and set into relation to the rank posed on the ligand samples by the *Diffdock* confidence score.

3.7 Molecular Dynamics Simulation

In order to validate the binding of the discovered compounds, we used *GROMACS* (Version 2023.3, Abraham et al. (2015)) to simulate the drugs inside the consensus 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 A-chain PDB file was used as input. As the ligands feature bonds and atoms not commonly seen in proteins, it is necessary to create a fitting force-field for them. For this, *acpype* (Silva and Vranken 2012) was used. The output files from *Maestro Glide* were converted to PDB and then piped into *acpype* which creates files that enable us to add the compound to the simulation environment. The output was then combined with the *GROMACS* topology and *.gro* files manually, according to Lemkul. Following the tutorial, the simulation space was set up and filled with water (a99SBdisp_water used as its force-field). Sodium and chloride ions were added to create a net-zero charge system. Energy-minimization and *NVT* / *NPT* equilibration were conducted. This ensures a somewhat relaxed starting point for the simulation. 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 GitHub repository. Additionally, a second MD simulation was run, using ANP as the ligand, in order to contextualize the result of the first simulation.
For the analysis of the simulation, *GROMACS* internal tools were used to calculate the RMSD and RMSF of the protein, according to the *GROMACS* manual (Abraham et al. 2015). 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* (Humphrey et al. 1996). For MM-PBSA calculations we used *gmx_MMPBSA* (Miller et al. 2012). This package enables direct estimation of binding affinities using the *GROMACS* output files.

4 Results

4.1 Consensus binding site detection

The first step of our drug discovery pipeline, is the detection of the consensus binding site using three distinct computational tools. Overall, the hot spots calculated with *Fpocket* overlapping between the three crystal structures all ranked first among the calculated binding sites for each crystal structure. Their DScores were 0.647 for 6ZSL, 0.383 for 5RM2, and 0.875 for 5RME. Also, the respective overlapping binding sites calculated with *PrankWeb* ranked either first (5RM2 and 5RME) or second (6ZSL) among the calculated binding sites. Therefore, the respective *Scores* and calculated *Probability* was also among the highest for these binding sites (Score of 9.69 and Probability of 0.562 for 6ZSL, Score of 22.15 and Probability of 0.860 for 5RM2, and a Score of 17.91 and a Probability of 0809 for 5RME). Additionally, four hot spots were detected for 6ZSL and 5RME which contained some amino acids in common with the just mentioned binding site. These were also merged into the consensus binding pocket. However, they were ranked lower by *PrankWeb* and also contained additional amino acids not present in the highest ranked binding sites. Interestingly, for the crystal structure 5RME a second binding hot spot was detected in proximity to the 1B domain. Nevertheless, because this binding site was only found in one crystal structure and the underlying pockets were ranked very low by both *Fpocket* and *PrankWeb*, this binding site was not considered for further analysis. The resulting consensus binding pocket is located in the ATP binding pocket in between

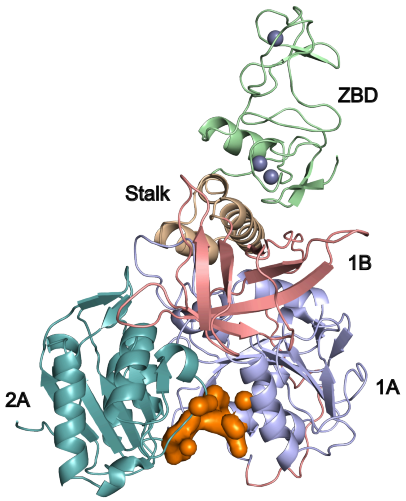


Figure (1) Consensus binding pocket of the NSP13 helicase. The binding pocket is located in the ATP binding pocket in between the two RecA-like domains A1 and A2.

the two RecA-like domains A1 and A2 (Figure 1).

4.2 Evaluation of Molecular Docking

The 100 top scorers from Vina were subjected to Glide. The lead compounds as per Glide are listed in Table 1. All of them had a higher docking score and therefore higher affinity to the receptor than ADP, but a lower affinity compared to ATP.
In comparison, Vina and Glide results showed no correlation between their ranks (Figure 2). The Pearson correlation coefficient r of the scores is 0.13 with a p-value of 0.19 which is similar to the r value of the ranks at 0.12 with a p-value of 0.22. However, 3 of Vina's top 5 scorers (rank 1, 2 and 4) were still part of the Top 10 Glide results (rank 9, 10 and 8).

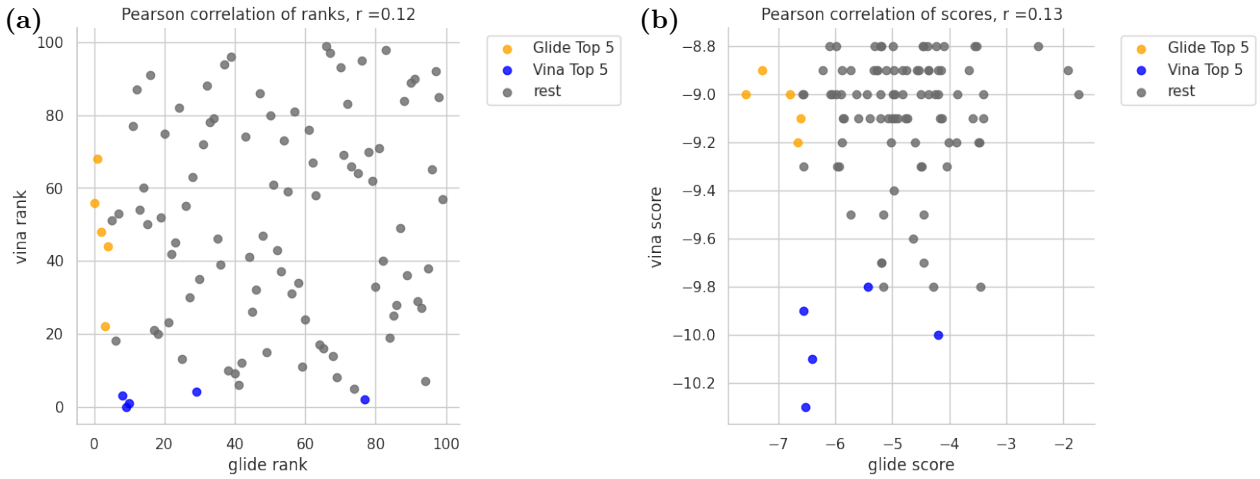


Figure (2) Comparison of Vina and Glide docking results. 100 ligands were compared in a) ranks and b) docking scores. Dots in blue represent the top 5 ligands from Vina and in orange the top 5 from Glide. All the other top 100 ligands are marked grey.

Analysis of intermolecular interactions between ligand and receptor binding pocket showed many bonds our Top 1 scorer from Glide has in common with ADP and ATP (LYS 288, LYS 320 and ARG 443). LYS 288 was thereby of particular interest, as it further interacts with our Top 2 and Top 3 scorer (Table 2). In most cases these bonds are either hydrogen bonds or salt bridges with some exceptions like an aromatic hydrogen

bonds, halogen bonds or pi-cation.

4.3 Validation of binding-site for the Top10 scorer of Glide

As shown in Figure 3, *Diffdock* did not validate our binding site for the Top 1 and Top 2 Glide results. None of the predicted conformations were found within our binding pocket. The Top 3 Glide result had two of *Diffdocks* best ranked conformations within our binding pocket. Furthermore, *Diffdocks* confidence into conformations of our Glide Top 3 ligand seems to correlate with their position within our binding pocket (Figure S5). This supports the assumption that the preferred binding site for this ligand is within our binding pocket. Among our Top 10 ligands from *Glide*, better results according to *Diffdock* were only achieved for our Top 5 ligand from *Glide*.

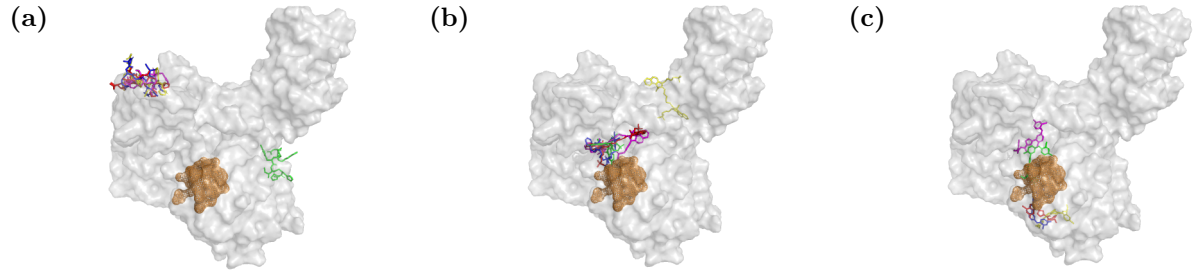


Figure (3) Diffdock results for our Top 3 Glide ligands. Shown are always the 5 ligand conformations Diffdock was most confident in. (a) Glide Top-1 (b) Glide Top-2 and (c) Glide Top-3.

4.4 Top 100 Compounds Exhibit Low Toxicity and High Synthetic Accessibility

After identifying the top 100 best-binding ligands through *AutoDock Vina*, our subsequent analysis focused on evaluating their practical applicability as potential drugs by considering their predicted toxicity and synthetic accessibility. The resulting SAscore and Tox-Score for each compound were visualized in a scatter plot as seen in Figure 4.
Of those 100 compounds, 99 presented with a Tox-score below the threshold of toxicity, indicating a low probability of being toxic to humans. The overall median Tox-Score is 0.24, with a mean of 0.26 across all compounds. Across all compounds the median SAscore of 2.69 and a mean of 2.87 which suggests that they are generally easy to synthesize. The top scorer from Glide, ZINC000096077632, was predicted to have a Tox-Score of 0.14 and an SAscore of 5.02.

4.5 Molecular Dynamics Simulation Validates Binding of Top Scorer

The MD simulation is integral to validate the results of our pipeline. After the production run was finished, the resulting trajectory file was centred on the protein and modified to remove any ghosting and splitting of the protein at the borders of the simulation box. The last frame of the simulation was extracted and visualized using *pyMOL* (Schrödinger, LLC 2015), which can be seen in Figure 5. The figure clearly demonstrated, that the top scoring compound of our Glide screening stayed inside the binding pocket until the end of the simulation. This was also confirmed by manual inspection of the frames generated in the simulation, which were combined to a video file using *VMD*. Furthermore, the polar interactions between the ligand and residues in its proximity are shown. It seems, that EOS100380 is bound to the protein by interactions with Glu319, Lys320 and Arg442 of the NSP13 helicase. These interactions could also be validated using *LigPlot+* (Laskowski and Swindells 2011).
To investigate the dynamics of these interactions between ligand and protein, the RMSD and RMSF were calculated, which can be observed in Figure 6. The RMSD observed over the course of the simulation (Figure 6A) rises at first, but reaches a plateau in both the ligand and the protein. The RMSD seems to settle faster in the ligand, which exhibits less movement in general. Investigation into the RMSF per

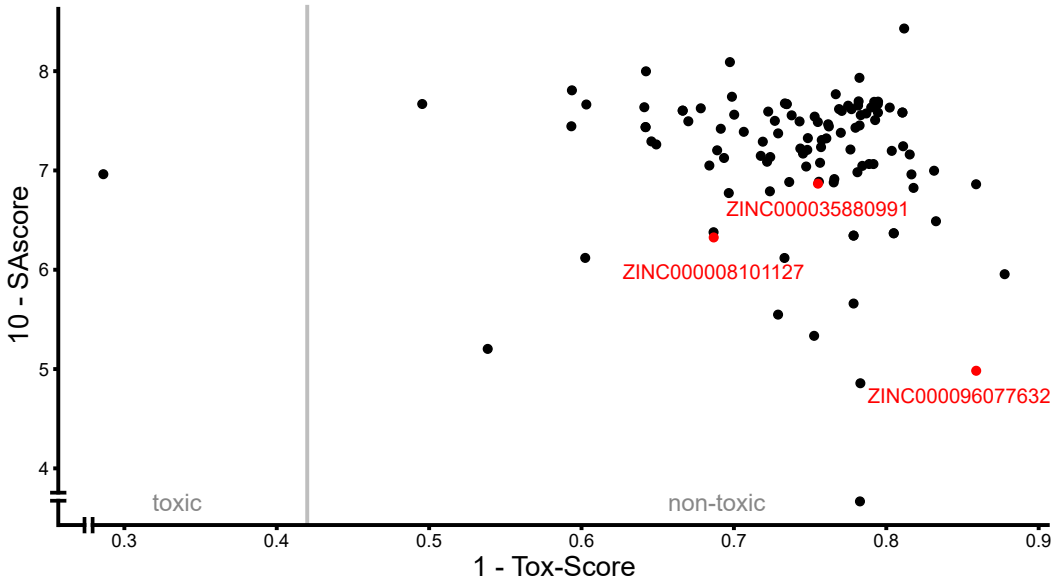


Figure (4) **Scatter plot of predicted toxicity and synthetic accessibility of the 100 best-binding compounds.** The predicted SAscore was plotted against the Tox-Score for the top 100 best scorers from *AutoDock Vina*. The top 3 scorers from *Glide* are highlighted in red. The vertical grey line represents the threshold for the toxicity, with the compounds to the right of this line being considered non-toxic.

residue (Figure 6B) suggests, that some residues move significantly more than others. The highly mobile residues seem to be located in the zinc binding domain and domains 1B and 2A. The latter two play a role in the binding of the ligand. Thus, movement in these parts of the protein was to be expected. All in all, the protein and ligand seems to undergo conformational changes, however the system does not show major signs of instability. The ligand seems to settle in a somewhat stable conformation after roughly 50ns. More information on the movement of the protein can be seen when investigating the radius of gyration (Supplementary Figure S3). The estimation of the interaction energy by MM-PBSA returned a $\Delta G_{solv} = -9.98 \frac{\text{kcal}}{\text{mol}}$. The RMSD and RMSF of the simulation using ANP as the ligand can be seen in Supplementary Figure S4. The latter system shows a generally lower RMSD and an earlier plateau compared to the first simulation. For the RMSF similarities between the runs can be observed, especially looking at the ZBD.

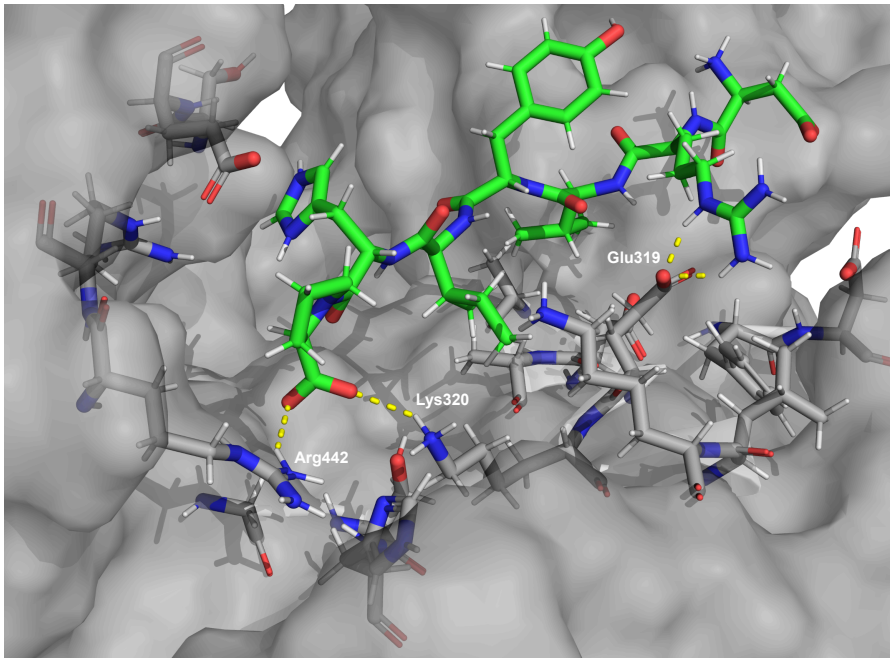


Figure (5) **Visualization of EOS100380 inside the binding pocket at the end of 100 ns simulation.** The polar interactions are marked with yellow dotted lines. The amino acids involved in these interactions are labelled.

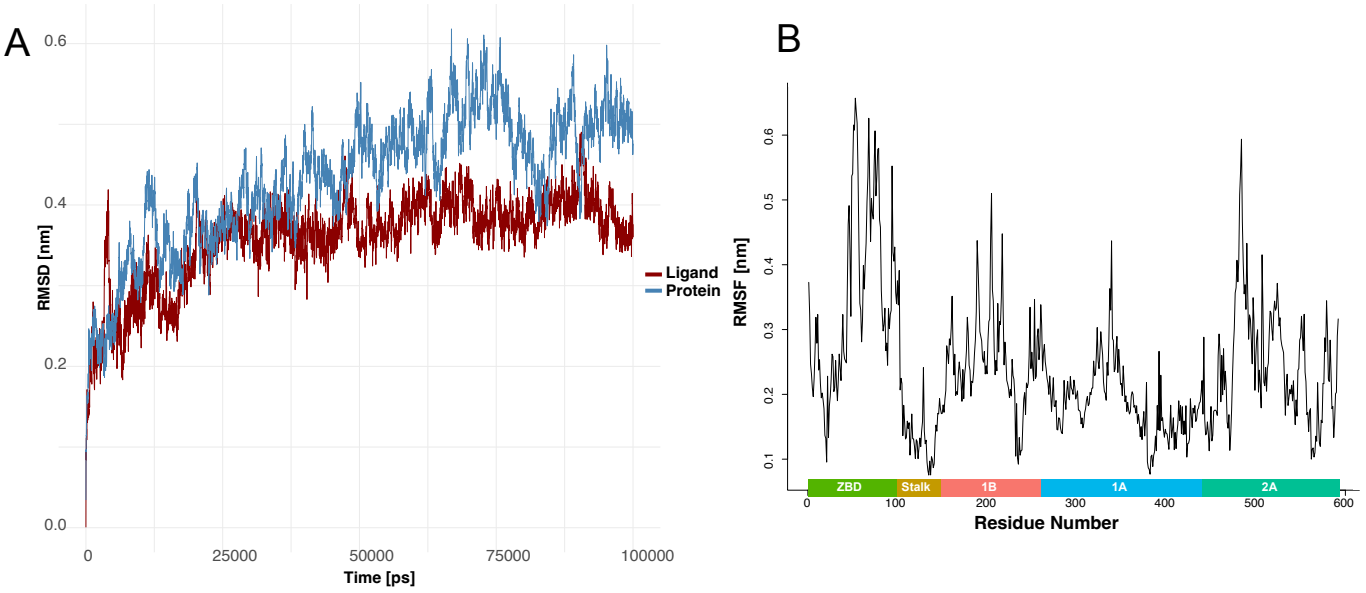


Figure (6) **RMSD of the protein backbone and ligand (A) and RMSF of the protein backbone per residue during the simulation (B).** The RMSD was plotted over time for the whole 100 ns simulation. The respective RMSDs are shown in red (ligand) and blue (protein) The RMSF is shown per residue in the residue. The colour bar annotates the protein domain the respective residue belongs to, according to Jia et al. (2019). ZBD = Zinc Binding Domain

5 Discussion and Outlook

In this work, we identified a consensus binding pocket of the NSP13 helicase using three distinct computational tools. It is located between the RecA-like domains A1 and A2 in the ATP binding pocket. This is in agreement with the work by Berta et al. (2021) who identified structurally important conserved motifs in the ATP binding site of SARS-CoV-2 NSP13. Out of the 16 amino acids they depicted to interact with ATP 13 are located in the consensus binding pocket identified by us. We conclude, that the combination of the three binding site detection tools *Fpocket*, *PrankWeb*, and *FTMap* is a suitable approach to identify a consensus binding pocket. However, the consensus binding pocket is not necessarily the best binding site for a potential inhibitor of the NSP13 helicase, since the natural ligand ATP is the primary energy currency of the human body, therefore it plays a crucial role in fundamental processes. Thus, inhibiting the ATP binding site selectively is challenging. If failed, this could lead to severe side effects. Therefore, it is important to consider other binding sites as well.

The Tox-Score predictor of *e*ToxPred was trained using FDA-approved dataset as non-toxic incidences. Consequently, the low mean Tox-Score of the tested compounds aligns with our expectation, considering that the compounds from the ZINC database are derived from an FDA-approved dataset. The single toxic incidence we detected was derived from the ECBD database which was consists of FDA-approved and non-FDA-approved molecules. As highlighted by Pu et al., natural compounds typically exhibit higher SAscore values compared to synthetic compounds due to their inherent complexity (Pu et al. 2019). The relatively high SAscore of ZINC000096077632 can be explained by the fact that ZINC000096077632 corresponds to angiotensin-(1-7) which is a naturally occurring compound with a crucial role in the renin-angiotensin system (RAS) (Santos 2014). The analysis of natural compound datasets by Pu et al. revealed a bimodal distribution in the SAscore, with peaks around 3 and 5. Furthermore, the very low Tox-Score of the top scorer can also be explained by the fact that it is a naturally occurring molecule in the human body.

The investigation into the molecular dynamics (MD) simulation further validated the viability of our pipeline. The RMSD of the backbone protein atoms (see Figure 6A) was slightly higher than in the tutorials provided by Lemkul (2018). However, the system stabilized over the duration of the simulation. Looking at the RMSF per residue (see Figure 6B) it is apparent, that certain domains of the protein lead to this increase in the RMSD. Especially the zinc binding domain shows a lot of movement. This was also observed throughout the simulation with ANP as the ligand, which serves as a positive control. Its RMSD was lower in general, which was to be expected as it is structurally very similar to ATP and the PDB file used was generated by crystallography of ANP bound to the protein, however the similarity gives confidence in the stability of our system. When discussing the general movement of domains in such simulations one should keep in mind, that the NSP13 A-chain is part of a much bigger replication complex (Newman et al. 2021) in an *in-vitro* setting. The conformational changes seen in this analysis would be severely hindered by other components in the complex. The domains 1B and 2A were suspected to be involved in a lot of conformational changes, as they make up a large part of the interaction surface between the protein and the ligand. The RMSD is calculated versus the output of Glide as a reference structure. It is completely possible that the system created by Glide is not completely relaxed. Thus, conformational changes are to be expected when protein and ligand can suddenly move freely. Investigating the RMSD of the ligand shows, that it too settles in a stable conformation during the simulation time span. Part of the RMSD can be attributed to functional groups far away from the binding site rotating along their sigma bounds, which was observed during manual inspection of the simulation. The result of the MM-PBSA calculation also suggests that strong interactions happen between the protein and our ligand, which can be seen by a negative ΔG_{solv}. All in all, the analysis offers confidence, that Angiotensin 1-7 binds to the NSP13 helicase and stays bound through an extended period of time. This fits the work of Mendoza-Torres et al. (2023), who propose Angiotensin 1-7 as a new therapy to support the recovery from Covid-19. Next to its role in the Renin-Angiotensin-Aldosteron System, it seemingly could also lower the viral damage to the body by inhibiting NSP13.

6 Limitations of the project

Regarding the molecular dynamics (MD) simulation, the project was held back by the tight time schedule. Running simulations for other top scoring ligands would have helped immensely in validating out pipeline further and compare the results of the top scorer. Furthermore, running multiple runs of the same simulation would have also further deepened our confidence in the results, as MD simulations in and of themselves are a very stochastic process and should always be estimated using replicas. Following the work of Raubenolt et al. (2022), implementing replica exchange with dynamical scaling would have elevated the validation step of this project. Access to more GPUs would have also made multiple simulations at once possible. For the analysis of the protein-ligand complex using *gmxMMPBSA* the same calculation should have been conducted on the MD run using ANP as the ligand. However, due to technical difficulties the calculation could not be performed in time. Thus, the given ΔG_{solv} of our ligand is a little difficult to interpret without a comparison value. Furthermore, the original plan to implement *AutoGrow4* (Spiegel and Durrant 2020) in order to improve our lead drugs and generate novel compounds was hindered by technical problems, which could not be fixed with the limited time at our disposal. We believe, that the increase in diversity among the drug candidates would lead to a better final drug candidate. With the help of *eToxPred*, as shown in this report, the new compounds could have been evaluated for their toxicity and a well binding, not too cytotoxic lead could have been presented.

7 Supplementary Material

Table (1) Glide docking scores for the top 4 ligands, ATP and ADP.

Title	docking score
ATP	-7.991
EOS100380	-7.573
ZINC000008101127	-7.273
EOS100897	-6.786
ZINC000150588351	-6.655
ADP	-6.63

Table (2) List of intermolecular interactions between ligands and receptor protein.

Ligand	Bond type	Amino acid
ADP	hydrogen bond	GLY 287, LYS 288, SER 289, HIE 290, LYS 320, ARG 443, SER 539
	aromatic h-bond	LYS 320
	salt bridge	LYS 288, LYS 320, ARG 443
	Pi-cation	ARG 443
ATP	hydrogen bond	GLY 287, LYS 288, SER 289, HIE 290, ARG 443, GLH 375
	salt bridge	LYS 288, LYS 320, ARG 443
EOS100380	hydrogen bond	LYS 288, LYS 320, ASP 315, GLU 319, GLU 341, ARG 443, GLU 540
	salt bridge	LYS 288, LYS 320, ASP 315, GLU 319, ARG 443, GLU 540
ZINC000008101127	hydrogen bond	ARG 178
	aromatic h-bond	GLY 538
	salt bridge	LYS 288
EOS100897	hydrogen bond	ASP 534
	halogen bond	ASN 516, THR 532
	Pi-cation	LYS 288

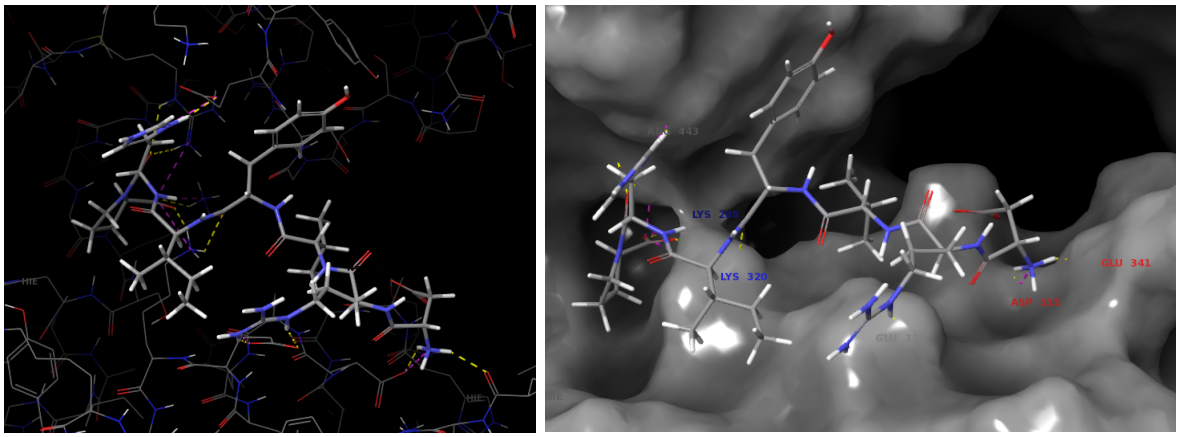


Figure (S1) Molecular docking of the top scoring ligand with Glide. The result of the docking process with the top scoring ligand (ECBD ID: EOS100380) is shown with focus on a) the specific interactions between ligand and receptor protein and b) position within the binding pocket. The different interactions between the two molecules are coloured depending on type. Hydrogen bonds are shown as yellow and salt bridges as violet.

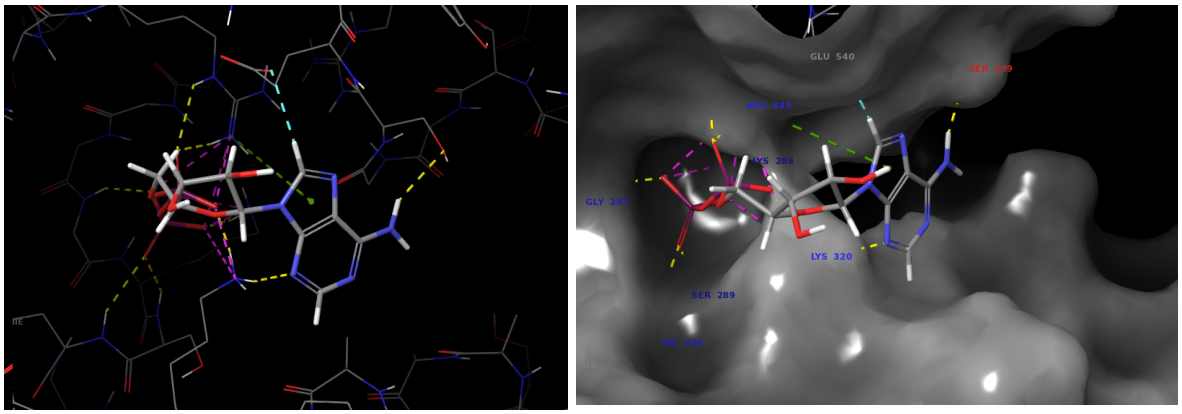


Figure (S2) Molecular docking of the ADP with Glide. The result of the docking process with ADP is shown with focus on a) the specific interactions between ligand and receptor protein and b) position within the binding pocket. The different interactions between the two molecules are coloured depending on type. Hydrogen bonds are shown as yellow, salt bridges as violet, aromatic hydrogen bonds as blue and pi-cation as green.

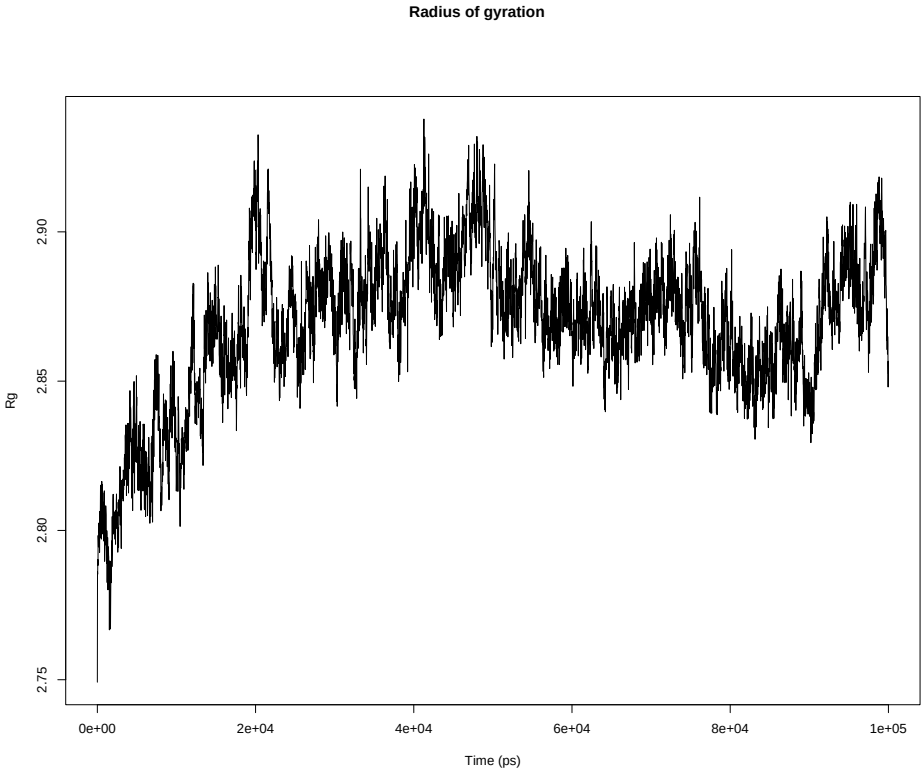


Figure (S3) Radius of Gyration of the NSP13 protein throught the production run featuring EOS100380.

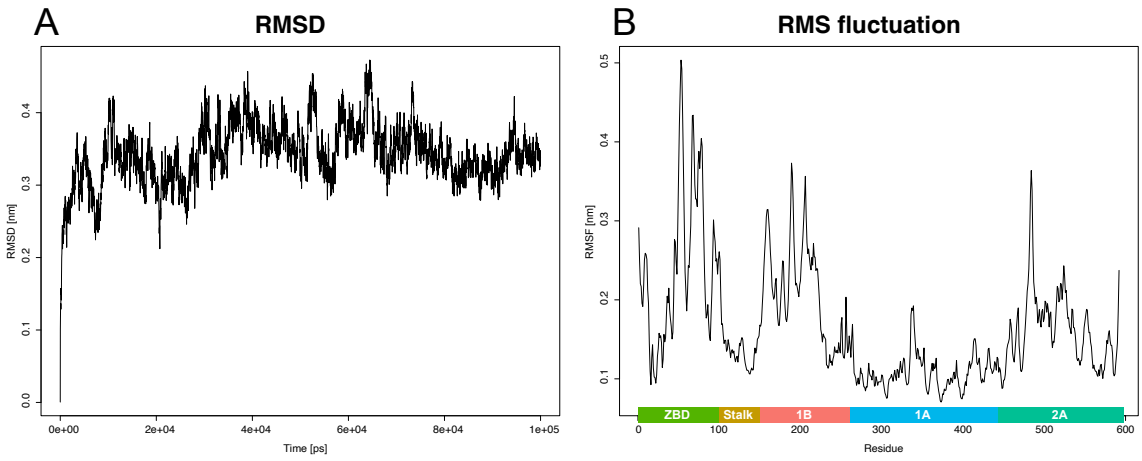


Figure (S4) RMSD and RMSF of Protein backbone for the simulation with ANP

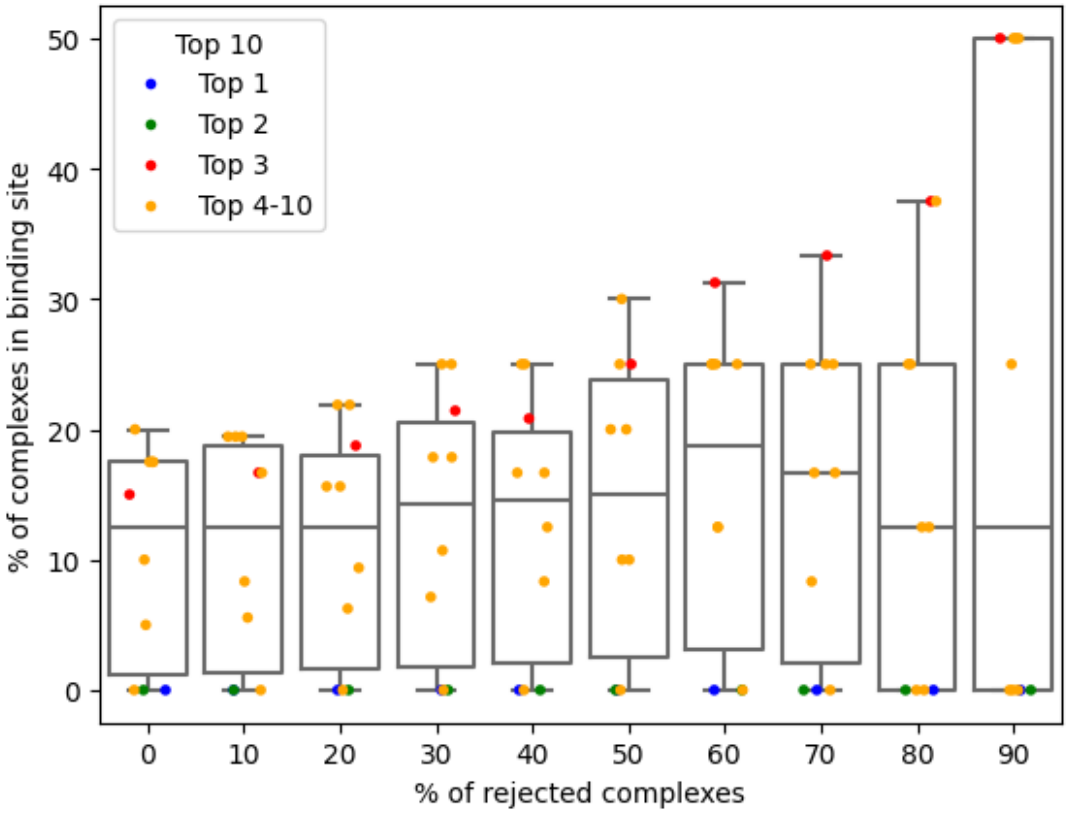


Figure (S5) Percentage of ligand conformation within our binding site in relation to the selection made by Diffdock

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.

Afgan, E., D. Baker, M. Van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, et al. (2016). "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update". In: *Nucleic acids research* 44.W1, W3–W10.

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.

Boda, K., T. Seidel, and J. Gasteiger (2007). "Structure and reaction based evaluation of synthetic accessibility". In: *Journal of computer-aided molecular design* 21, pp. 311–325.

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.

Che, X., Q. Liu, and L. Zhang (2023). "An accurate and universal protein-small molecule batch docking solution using Autodock Vina". In: *Results in Engineering* 19, p. 101335. ISSN: 2590-1230. DOI: https://doi.org/10.1016/j.rineng.2023.101335. URL: https://www.sciencedirect.com/science/article/pii/S2590123023004620.

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.

Dearden, J. C. (2003). "In silico prediction of drug toxicity". In: *Journal of computer-aided molecular design* 17.2-4, pp. 119–127.

Durrant, J. D. and J. A. McCammon (2011). "Molecular dynamics simulations and drug discovery". In: *BMC Biology* 9.1, p. 71. DOI: 10.1186/1741-7007-9-71.

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.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin (2004). "Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy". In: *Journal of Medicinal Chemistry* 47.7. doi: 10.1021/jm0306430, pp. 1739–1749. ISSN: 0022-2623. DOI: 10.1021/jm0306430. URL: https://doi.org/10.1021/jm0306430.

Genheden, S. and U. Ryde (2015). "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities". In: *Expert Opinion on Drug Discovery* 10.5, pp. 449–461. ISSN: 1746-0441. DOI: 10.1517/17460441.2015.1032936.

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.

Humphrey, W., A. Dalke, and K. Schulten (1996). "VMD – Visual Molecular Dynamics". In: *Journal of Molecular Graphics* 14, pp. 33–38.

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 Dis-

covery". 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.

Jia, Z., L. Yan, Z. Ren, L. Wu, J. Wang, J. Guo, L. Zheng, Z. Ming, L. Zhang, Z. Lou, and Z. Rao (May 2019). "Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis". In: *Nucleic Acids Research* 47.12, pp. 6538–6550. ISSN: 0305-1048. DOI: 10.1093/nar/gkz409. eprint: https://academic.oup.com/nar/article-pdf/47/12/6538/28917080/gkz409.pdf. URL: https://doi.org/10.1093/nar/gkz409.

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.

Laskowski, R. A. and M. B. Swindells (2011). "LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery". In: *Journal of Chemical Information and Modeling* 51.10, pp. 2778–2786. ISSN: 1549-9596. DOI: 10.1021/ci200227u.

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.

Lu, C., C. Wu, D. Ghoreishi, W. Chen, L. Wang, W. Damm, G. A. Ross, M. K. Dahlgren, E. Russell, C. D. Von Bargen, R. Abel, R. A. Friesner, and E. D. Harder (2021). "OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space". In: *Journal of Chemical Theory and Computation* 17.7. doi: 10.1021/acs.jctc.1c00302, pp. 4291–4300. ISSN: 1549-9618. DOI: 10.1021/acs.jctc.1c00302. URL: https://doi.org/10.1021/acs.jctc.1c00302.

Madhavi Sastry, G., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman (2013). "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments". In: *Journal of Computer-Aided Molecular Design* 27.3, pp. 221–234. ISSN: 1573-4951. DOI: 10.1007/s10822-013-9644-8. URL: https://doi.org/10.1007/s10822-013-9644-8.

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.

Martinez, L. (2015). "Automatic Identification of Mobile and Rigid Substructures in Molecular Dynamics Simulations and Fractional Structural Fluctuation Analysis". In: *PLoS ONE* 10.3, e0119264. DOI: 10.1371/journal.pone.0119264.

Mendoza-Torres, E., P. Luna, M. Fernanda Pérez, J. Castellar-Lopez, A. Chang, Y. Montoya, J. Bustamante, and W. Rosales-Rada (Jan. 2023). "Potential of Angiotensin-(1-7) in COVID-19 Treatment". In: *Current Protein & Peptide Science* 24.1, pp. 89–97. ISSN: 1389-2037. DOI: 10.2174/1389203724666221130140416.

Miller, B. R. 3., T. D. J. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg (Sept. 2012). "MMPBSA.py: An Efficient Program for End-State Free Energy Calculations." eng. In: *Journal of chemical theory and computation* 8.9, pp. 3314–3321. ISSN: 1549-9618 (Print). DOI: 10.1021/ct300418h.

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.

O'Boyle, N. M., M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison (2011). "Open Babel: An open chemical toolbox". In: *Journal of cheminformatics* 3.1, pp. 1–14.

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.

Raies, A. B. and V. B. Bajic (2016). "In silico toxicology: computational methods for the prediction of chemical toxicity". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 6.2, pp. 147–172.

Raubenolt, B. A., N. N. Islam, C. M. Summa, and S. W. Rick (2022). "Molecular dynamics simulations of the flexibility and inhibition of SARS-CoV-2 NSP 13 helicase". In: *Journal of Molecular Graphics and Modelling* 112, p. 108122. ISSN: 1093-3263. DOI: 10.1016/j.jmgm.2022.108122.

Ravindranath, P. A., S. Forli, D. S. Goodsell, A. J. Olson, and M. F. Sanner (2015). "AutoDockFR: advances in protein-ligand docking with explicitly specified binding site flexibility". In: *PLoS computational biology* 11.12, e1004586.

Ricci, F., R. Gitto, G. Pitasi, and L. De Luca (2022). "In Silico Insights towards the Identification of SARS-CoV-2 NSP13 Helicase Druggable Pockets". In: *Biomolecules* 12.4. 2218-273x Ricci, Federico Gitto, Rosaria Pitasi, Giovanna Orcid: 0000-0003-4694-9070 De Luca, Laura Orcid: 0000-0003-0614-5713 Journal Article Switzerland 2022/04/24 Biomolecules. 2022 Mar 22;12(4):482. doi: 10.3390/biom12040482. ISSN: 2218-273x. DOI: 10.3390/biom12040482.

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.

Roncaglioni, A., A. A. Toropov, A. P. Toropova, and E. Benfenati (2013). "In silico methods to predict drug toxicity". In: *Current opinion in pharmacology* 13.5, pp. 802–806.

Santos, R. A. (2014). "Angiotensin-(1–7)". In: *Hypertension* 63.6, pp. 1138–1147.

Schrödinger Release 2022-3: Maestro. Schrödinger, LLC; New York, NY, USA: 2021. (2021).

Schrödinger, LLC (Nov. 2015). "The PyMOL Molecular Graphics System, Version 1.8".

Schrödinger, LLC (2023). "The PyMOL Molecular Graphics System, Verion 2.5.7". available online. URL: httpps://pymol.org (visited on 12/27/2024).

Silva, A. W. S. d. and W. F. Vranken (2012). "ACPYPE - AnteChamber PYthon Parser interfacE". In: *BMC Research Notes* 5.1, p. 367. DOI: 10.1186/1756-0500-5-367.

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.

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.

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.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN: 1441412697.