

European Bioinformatics Masters Network

Meet EU



CHARLES UNIVERSITY

Prague

Authors:

Arnošt Polák, Jan Faflík, Radoslav Jochman, Sofya Safiulina, Adam Kuniak

Advisors:

Marian Novotný, Ph.d., Rafael Najmanovich, Ph.d., Jiří Černý, Ph. d, Miroslav Lžičar

Partner team:

Sorbonne 2

Table of contents

1. Introduction	3
1.1. Problem Statement	3
1.2. Nsp-13	3
2. Methods	3
2.1. Overall approach	3
2.2. Virtual screening methods.....	4
2.3. Choice of protein structures	4
2.4. Pipeline Description	5
3. Results.....	6
3.1. Top 100 hits from the Autodock vina on 7RDY	6
3.2. Conformation analysis with Autodock Vina	8
3.3. DiffDock results	9
3.4. NRGDock results	9
4. Discussion	10
5. References.....	12

1. Introduction

1.1. Problem statement

The emergence of the novel coronavirus, SARS-CoV-2, and the subsequent global pandemic of COVID-19 has presented an unprecedented challenge to public health and the global community. The pandemic, as of the end of last year, has tragically resulted in over 7 million deaths worldwide (WHO, 2024), with many more suffering from the repercussions of the virus. Either in terms of health or economic impact.

Despite a decrease in mortality rates, the virus continues to impact many people around the world. In lieu of this, there is a need to understand better the virus, how it operates, and how to target it to stop its spread and subsequent effects. That is our goal in this project. We used in-silico high throughput screening methods to identify new molecules inhibiting the virus replication process. To achieve that, in our screens, we targeted a protein known for its vital importance in SARS-CoV-2 virus activity, non-structural protein 13 (Nsp-13), and identified multiple promising inhibitor candidates.

1.2. Nsp-13

Nsp13 is a 67 kDa protein that belongs to the helicase superfamily 1B. It utilizes the energy of nucleotide triphosphate hydrolysis to catalyze the unwinding of double-stranded DNA or RNA in a 5' to 3' direction (Tanner *et al.*, 2003). Although Nsp13 is believed to act on RNA in vivo, enzymatic characterization shows a significantly more robust activity on DNA in vitro assays with relatively weak non-processive helicase activity compared to other superfamily 1B enzymes (Jang *et al.*, 2020). Nsp13 has been shown to interact with the viral RNA-dependent RNA polymerase NSP12 (Jia, Z. *et al.*, 2019) and acts in concert with the replication-transcription complex (NSP7/NSP8/NSP12) (Chen, J. *et al.*, 2020). This interaction has been found to significantly stimulate the helicase activity of Nsp13, possibly by means of mechano-regulation (Adedeji *et al.*, 2012). In addition to its helicase activity, Nsp13 also possesses RNA 5' triphosphatase activity within the same active site (Ivanov *et al.*, 2004), suggesting a further essential role for Nsp13 in the formation of the viral 5' mRNA cap.

Nsp13 contains five domains: an N-terminal Zinc binding domain (ZBD) that coordinates three structural Zinc ions, a helical “stalk” domain, a beta-barrel 1B domain, and two helicase subdomains 1A and 2A that contain the residues responsible for nucleotide binding and hydrolysis. Nsp-13 has many binding sites that could potentially be targets for inhibitors. Former analyses of Nsp-13 have identified four pockets with a high likelihood of binding inhibitors. Pocket 1 plays a crucial role in the translocation of genetic material since the interaction, hydrolysis, and release of ATP triggers conformational changes in the enzyme and remodels the interface of the RNA. Pocket 2 forms the binding channel for the RNA/DNA strand. Pocket 3 is situated between the n-terminal portion involved in zinc binding (ZBD) and the stalk domain. Pocket 3 significantly contributes to the enzymatic activity of SARS-CoV-2 Nsp13. Deletion of the stalk domain or insertion of a linker between the stalk and helicase domains adversely affects the ATPase activity and stability of the enzyme. Pocket 4 has been identified at the interface between the zinc-binding domain (ZBD) and non-structural protein 8 (Nsp8). Pocket 4 plays a crucial role in forming a stable complex responsible for generating backtracked RTC. This complex is vital for proofreading and template switching during sub-genomic RNA transcription (Isabella Romeo *et al.*, 2022).

2. Methods

2.1. Overall approach

This research aims to find a molecule that can bind and inhibit the Nsp13 helicase. To do so, we chose to test and compare three independent virtual screening methods and dock a diverse set of ligands, including FDA-approved drugs, which are the main focus of this analysis. Because of that, if we had to describe the approach in one sentence, we would probably describe it as an **Identification of repurposable compounds by comparative analysis of classical and modern AI-based virtual screening methods.**

Drug repurposing is an approach to drug discovery that leverages compounds already clinically tested for the treatment of one set of indications to find potential drug candidates for another set of indications. A lot more is known about these drugs, such as their ADMET profiles. Therefore, the time of their approval process is significantly shortened as many of the clinical trials have already been done. Further, this approach provides a simple way of significantly decreasing the molecular search space that could be screened to a more manageable amount to a set of compounds known to be usable for the treatment of human diseases.

For those reasons, we chose one of the datasets available through the Academic download in the DrugBank database as our virtual screening starting point. This dataset contains 11,586 FDA-approved drugs. (Knox *et al.*, 2024)

2.2. Virtual screening methods

We have chosen to conduct our virtual screening using three independent molecular docking methods described below. Molecular docking is an *in-silico* approach to finding which ligands are better at binding to a given target. The methods we chose are AutoDock Vina, DiffDock, and NRGDock.

AutoDock Vina is an industry-standard docking method that can generate a 3D conformer of a given ligand in a given search box (user-determined binding site) with a predicted level of affinity to that place on the target protein's surface. It achieves that by turnkey computational docking based on a simple scoring function and rapid gradient-optimization conformational search. This allows it to be used as an efficient screening method. (Eberhardt et al., 2021) DiffDock is a relatively new AI-based molecular docking method. It is a diffusion generative model over the non-Euclidean manifold of ligand poses. It comes trained and prepared with a Confidence Model to rank generated poses based on the likelihood they are below 2 Å RMSD. For clarification, it does not accept a binding site and only outputs a score list of conformers for a given ligand. (Corso et al., 2022) We were also offered the opportunity to use a novel docking method that was developed by the Najmanovich Research Group (NRGDock), which accepts similar input to AutoDock Vina (target, ligand, binding site), however, it is able to screen 48.2 million molecules in one day (on a modern laptop). The exact inner workings of the method will be published later this year. To demonstrate its capabilities, this method was used on a different ligand dataset than the previous two methods - namely, the Enamine REAL Diversity Set.

2.3. Choice of protein structures

In deciding which protein structures to use for the screen, we tried to select the newest structures and those with the best resolution for structural alignment. Based on that and the available literature on Nsp13 helicase conformations, we have decided to focus our project on the below-mentioned protein structures (Chen et al., 2022). Table 1 summarizes the properties of each conformation that were important to us in our selection.

Table 1. Properties of the Nsp13 helicase conformations

Apo (Pdb ID: 7RDZ)	Engaged (Pdb ID: 7RDY)	Swiveled (Pdb ID: 7RE0)	1B-open (Pdb ID: 7RDX)
The Nsp13T RecA2 domain is rotated open by 21° with respect to RecA1 The RNA-binding site is empty Occupancy of the Nsp13T nucleotide-binding site is ambiguous	Closed onto each other more than the other structures and are thereby engaged most tightly with the RNA The RecA domains are 'closed' and grasp the downstream t-RNA single-stranded 5' segment emerging from the RdRp active site	The Nsp13T subunit swivels as a whole by 38° away from Nsp13F	The Nsp13T domain 1B is rotated 85° away from the Nsp13T RNA-binding channel

Only one of the previously mentioned conformations of the helicase is able to grasp the downstream t-RNA - the *engaged* conformation; because of that, we mainly focused on it in our project.

The process of preparation of the structures for the screening was straightforward. Firstly, Nsp13 in all four conformations was separated from the replication-transcription complex (RTC) in PyMol, and each pair of these conformations was aligned to each other, as seen in Figure 1. to compare the difference between the structures.

Subsequently, using PrankWeb (Krivák & Hoksza, 2018), binding sites for all conformations were identified, and it was revealed that there were significant similarities between the top-ranked pockets on all of the conformations.

Because past research has already explored two of the best-scored binding sites (Piplani et al., 2022), we decided to focus on the two next-best ones, which are located between the RecA1 and RecA2 domains of both promoters (Nsp13T and Nsp13F) and their positions in the domains are near-symmetrical, as can be seen in Figure 2.

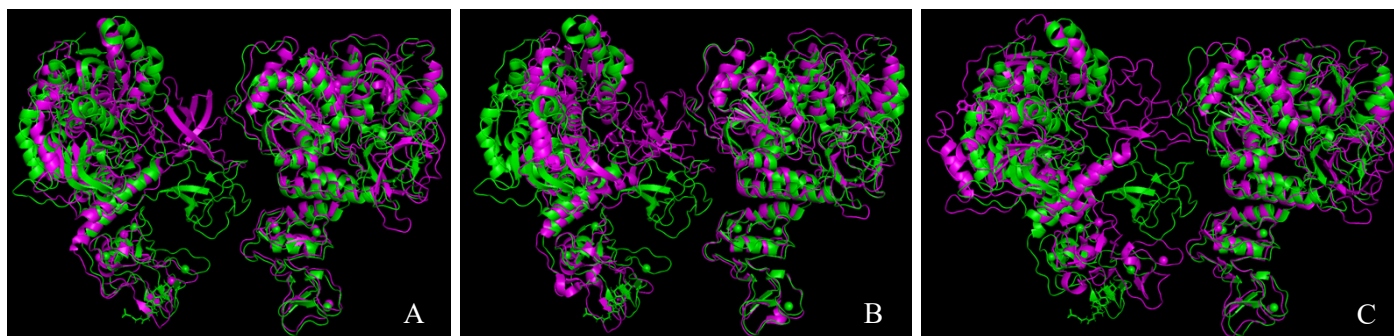


Figure 1. Structural alignment of Nsp13 structures in different conformations. In all of the pictures, the structures are aligned to 7RDX. A) alignment of 7RDY B) alignment of 7RDZ C) alignment of 7RE0

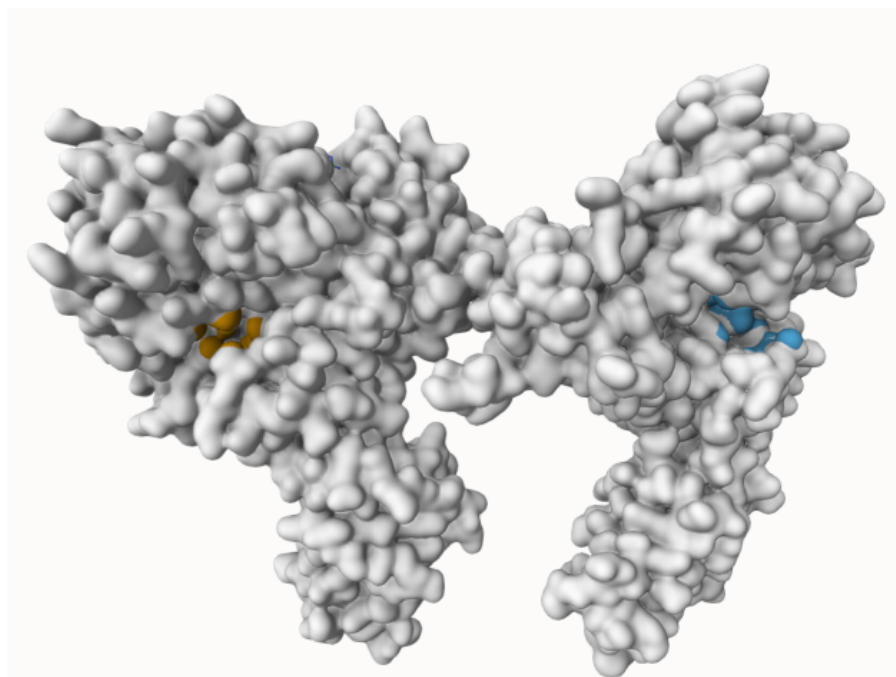


Figure 2. Selected pockets. The one on the left is located on Nsp13F, and the right one on Nsp13T. The 3rd top-ranked pocket is orange on the figure, and the 4th one is blue.

Subsequently, we created two versions of all helicase conformations for docking to see if already bound ligands from the solved structure will affect overall affinity. The first version had everything deleted apart from the actual protein (zinc atoms, phosphates, and other ligands). The second version only kept the zinc atoms.

All in all, we had four different protein targets with two different versions for docking with two predicted binding sites each to dock to.

2.4. Pipeline description

The pipeline can be deconstructed into two separate pathways. The first one uses AutoDock Vina and DiffDock. The second uses NRGDock, developed by the Najmanovich Research Group and AutoDock Vina.

The first pathway uses the DrugBank database of FDA-approved drugs. The molecules downloaded from the DrugBank site were in .sdf format, which needed to be converted to the .pdbqt format, which is accepted by AutoDock Vina. The conversion was done with OpenBabel. (O’Boyle et al., 2011) The proteins also needed to be converted from .pdb to .pdbqt. This conversion was done through the AutoDock Tools application, where any excess waters were deleted, non-polar hydrogens were added, and Kolman charges were added to the structure.

We identified that an average run of AutoDock Vina takes about 20 seconds. An average run of DiffDock takes 200 seconds. This resulted in a realization that running all the imagined combinations of proteins and molecules would either take too much time or too much computing power - the DrugBank database has over 11,500 molecules. Further, AutoDock Vina does not always produce consistent docking results. Therefore, we decided that each of the configuration combinations will be run ten times to produce a more consistent score and that one run of AutoDock Vina should not take more than 2 minutes, when that happened a hard cutoff was enforced and the ligand in question was skipped. All of the AutoDock runs were done on the Metacentrum clusters.

Therefore, the first step was to pre-screen the whole DrugBank database on one protein which would best represent the most applicable conformation to real-life scenarios. We decided to use the *engaged* conformation (protein structure 7RDY) to pre-screen all of the molecules from the DrugBank database and create a list of 100 best-scoring molecules. The screening process was done through both of the above identified binding pockets. With the information about their position and about their size the search box of AutoDock Vina was set to the center of the pocket and a cube of 20Å. Because of the ten repetitions of the runs, we only extracted the best binding affinity for each of the molecules. The rest of the configurations were left default.

These molecules were then redocked through DiffDock on the 7RDY structure. As DiffDock does not accept any binding site information, the identified sites were ignored. They were also screened again with AutoDock Vina on the rest of the conformations (protein structures 7RDX, 7RDZ, 7RE0). Each protein had their own binding pockets that had the same overall function but differed in their exact position on the protein (see section 2.3). The rest of the configurations were the same as before.

The second pathway uses NRGDock. The docking targeted the two identified binding sites of the 7RDY protein. The procedure was done by the Najmanovich Research Group in Toronto on their computing cluster. The result of this method was a list from the best molecules that the method identified. We randomly choose a 1000 molecules subset from the best 5000 hits, identified by NRGDock, to be re-screened through AutoDock Vina on 7RDY, so we could compare the accuracy of the 2 methods.

For a more schematic overview of the pipeline, please see the image below (Fig. 3).

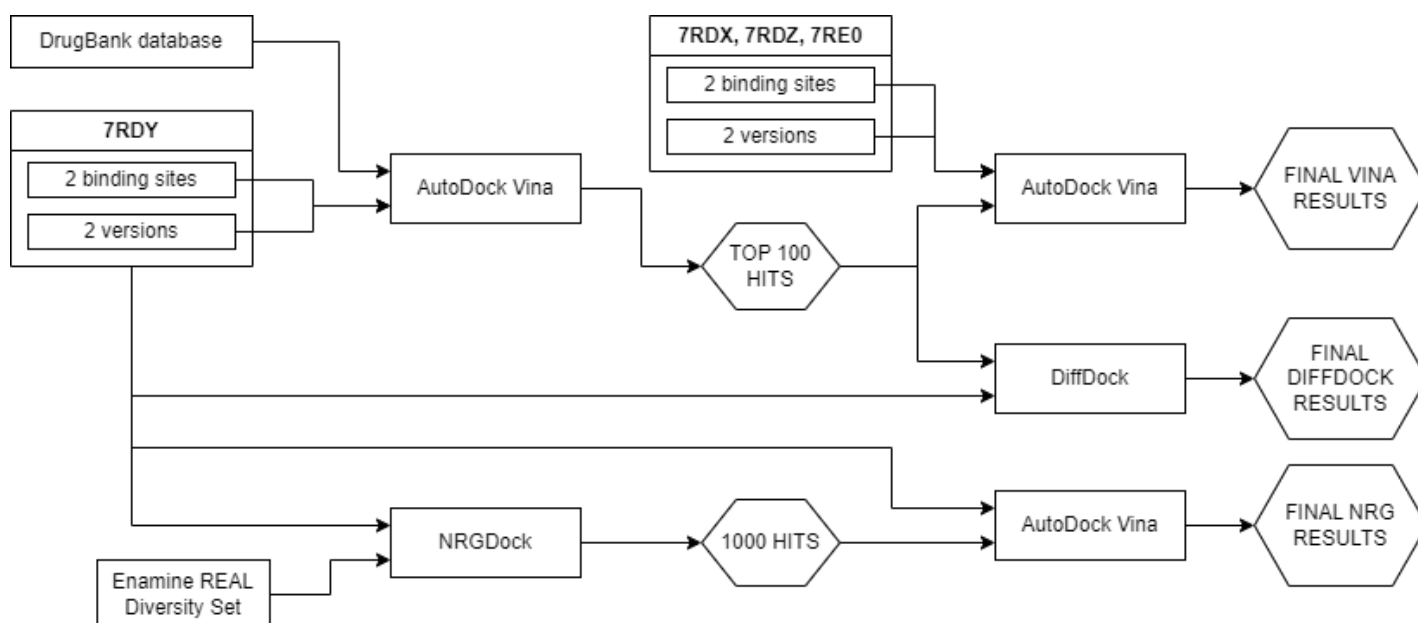


Figure 3. Pipeline schematic.

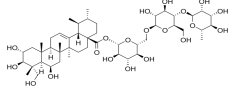
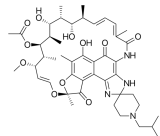
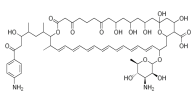
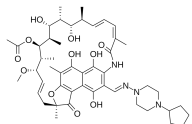
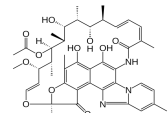
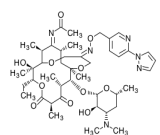
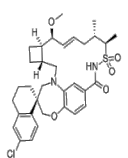
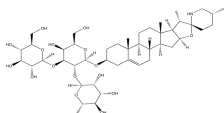
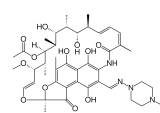
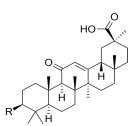
3. Results

As described above we have got 4 sets of results that were further processed and analyzed so we could gain more insights into the problem area. The sets of results are the Top 100 hits from the Autodock Vina on 7RDY screen, Conformation analysis with Autodock Vina (previous vina results redocked on 7RDX, 7RDZ, 7RE0), DiffDock results (on 7RDY) and NRGDock results (screened by both NRGDock and Autodock Vina on 7RDY).

3.1. Top 100 hits from the Autodock vina on 7RDY

The top 100 hits from the first AutoDock screen are characterized by a fairly large molecular weight (mean 659.37 g/mole) and a higher-than-average atom count (mean 94 atoms) in comparison with most drugs (Khanna & Ranganathan, 2009). The ten molecules with the best helicase binding affinity are summarized in Table 2. They can be generally characterized by neutral physiological charge, high number of rings in their structure (6 on average), high acceptor count (13 on average) and by generally not adhering to Lipinski's rule of five.

Table 2. Binding affinity of top 10 hits against SARS-CoV-2 helicase, from 1st Autodock vina screen

DrugBank ID	Drug Name	Drug Class	Structure	Autodock binding affinity (kcal/mol)
DB15532	Madecassoside	Anti-inflammatory agent		-17.75
DB00615	Rifabutin	Antibiotic		-15.85
DB01152	Candididin	Antimycotic		-15.50
DB01201	Rifapentine	Antibiotic		-14.80
DB01220	Rifaximin	Antibiotic		-14.77
DB16238	Modithromycin	Antibiotic		-14.77
DB17166	Tapotoclax	Anticancer agent		-14.55
DB17103	Solasonine	Anticancer agent		-14.51
DB01045	Rifampicin	Antibiotic		-14.48
DB13089	Enoxolone	Anti-inflammatory agent		-14.45

As said above the top 10 ten molecules share the same apparent characteristics and come mainly from 3 drug classes (Antibiotic, Anti-inflammatory and Anticancer agents) that are surprisingly not really related to virus activity inhibiting drugs.

3.2. Conformation analysis with Autodock Vina

The top 100 molecules from the 7RDY Autodock Vina screen were then redocked, into the rest of the Nsp13 helicase conformations (7RDX, 7RDZ, 7RE0). The data from the redocking were then processed, cleaned and analyzed. Because of the time restrictions the docking process had a fairly high error rate and after the cleaning and processing of the data we ended up with a dataset of 67 molecules that we docked to all of the structures.

Once processed the binding affinities data presented a subtle trend, where the molecules bound on average most to the 7RDX protein structure (open confirmation), then the 7RDY structure (engaged conformation), 7RDY (apo conformation) and least to the 7RE0 structure (swiveled confirmation). The details of the trend are summarized in Figure 4. Even though the experimental data showed that the 7RDX structure had the best ability to bind our ligand set, we decided that due to the reasons described in chapter 2.3, we will use the ligand list scored for the 7RDY binding affinity, as the starting point for further method comparisons.

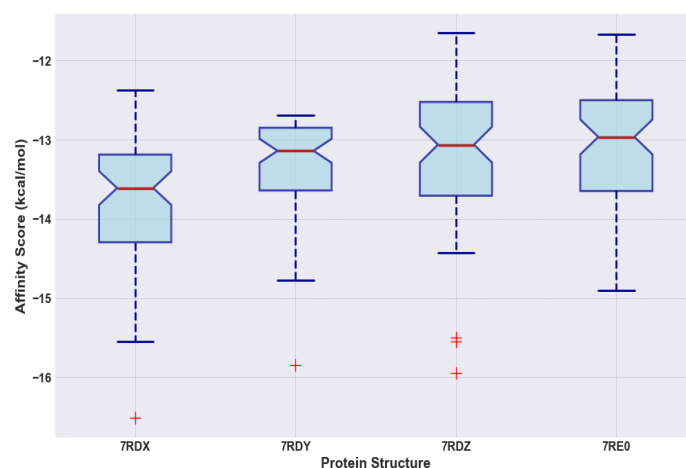


Figure 4. Box plot of the affinity scores of the top 100 ligands for the 4 conformations

Following 2 figures further demonstrate characteristics of the analyzed data. The protein structure with the highest number of minimal affinities (for given ligand, over the 4 structures), is 7RDX, while the protein structure with the highest number of minimal ranks (positions of given ligand in the scored list for given structure) is 7RDY. The details of the counts of minimal and maximal binding affinities and ranks are clearly summarized in Figure 5.

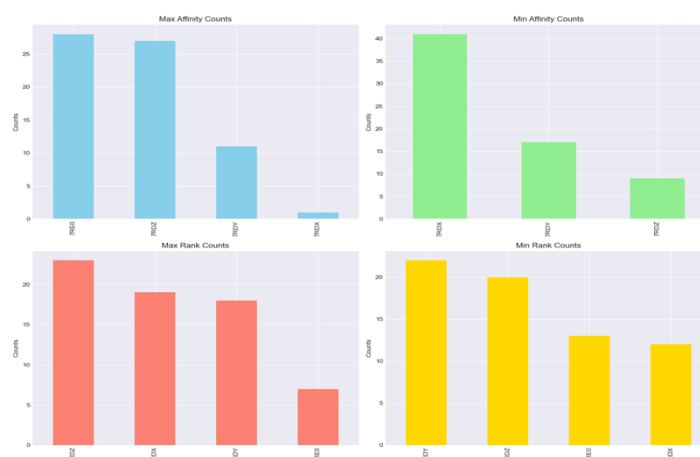


Figure 5. Counts of minimal and maximal binding affinities and ranks over the 4 structures.

Another interesting fact about the results, is that the variance in ranking over the 4 structures for given ligand, is in most cases not-insignificant for our dataset (mean var: 112), even when the distribution of the differences is significantly right-skewed. The details of the distribution of the rank variance are summarized in Figure 6.

This fact further demonstrates that even though there is a subtle trend in the binding affinities, the ranking of the best ligand with the dataset is fairly variable and not consistent over the 4 conformations. This possibly reflects the behavior of the Nsp13 helicase, where some conformations of the molecule clearly have a higher ability to bind outside ligands. Further, the specifics of the binding sites possibly change significantly with the conformation which has a clear impact on the types of ligands the molecule has high probability to bind.

3.3. DiffDock results

The top 100 molecules from the 7RDY Autodock Vina screen were redocked, by DiffDock, on the 7RDY structure. The data from the redocking were then processed, cleaned and analyzed. Because of the time restrictions we set to the docking process and the fact that DiffDock was not capable of generating RDKit coordinates consistently it had a fairly high error rate and after the cleaning and processing of the data we ended up with a dataset of 55 ligands, with calculated DiffDock confidence scores. DiffDock generated 40 conformers and predicted a confidence score for each of the ligands.

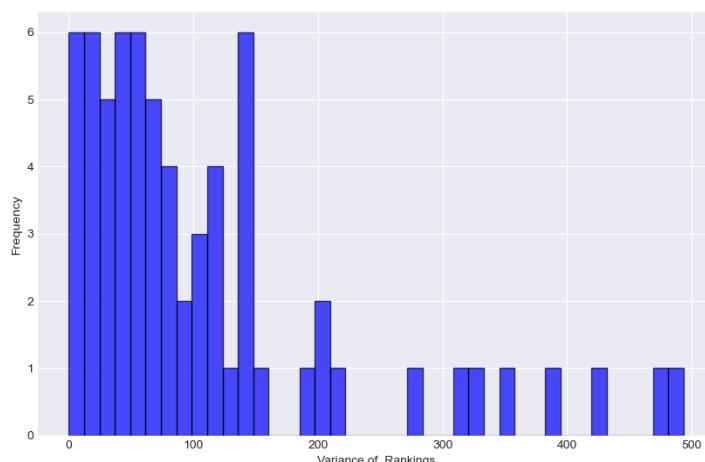


Figure 6. Distribution in the variance of ranking for the best 100 ligands docked in the 4 structures.

We tested 2 ranking methods on the DiffDock confidence scores, because we anecdotally saw that conformers that DiffDock scores the highest are many times placed on nonsensical location within the protein structure, however, over the 40 generated conformers Diffdock usually generates, most of the conformers were in proximity to a sensible binding site on the protein structure. For that reason, we used both the best confidence score DiffDock produced for given ligand and the mean confidence score DiffDock produced for given ligand as the ranking methods. Surprisingly the best confidence score method produced a slightly better mean absolute difference in between the DiffDock and Autodock Vina ranking (mean: 16.9) then the mean confidence score ranking method (mean: 18.1). The best confidence score method also produced a slightly better correlation to the Autodock Vina rank (Spearman's correlation coefficient: 0.084) than the mean confidence score ranking method (Spearman's correlation coefficient: 0.025).

From those results we can safely conclude that at least on the scale of less than 100 molecules there is no correlation between Autodock Vina scoring and DiffDock scoring.

3.4. NRGDock results

As said in the methods chapter, NRGDock was used to screen the Enamine REAL Diversity set (48.2m compounds) on the 7RDY structure. We then redocked a random 1000 ligand subset, from the best 5000 hits with Autodock Vina for comparison of the scoring. After the screening we processed and cleaned the results and tested them with 3-part analysis.

At first, we compared the structural similarity of the 1000 ligands by Morgan and MACCS fingerprints. MACCS fingerprints produced non-negligible similarity for more of the ligand pairs and therefore were determined to perform better on this dataset for our analysis. Overall, we determined the subset of molecules to be fairly diverse, as determined by the larger Enamine REAL Diversity criteria. For a closer look on the similarities between the ligands see Figure 7.

Secondly, we ranked the dataset by both the NRGDock and Autodock Vina generated scores and calculated the absolute difference in between the 2 ranks for each of the ligands.

The mean absolute difference between the NRGDock and Autodock Vina ranks for each of the ligands was approximately 335, which is quite a high number considering the rescored subset was 1000 ligands in size. We also calculated a Spearman's correlation coefficient 0.005 for the NRGDock and Autodock Vina ranking. From those results we can safely conclude that at least on the scale of less than a 1000 molecules there is no correlation between Autodock vina scoring and NRGDock scoring, which isn't necessarily a problem with a NRGDock since it is mainly designed to reduce data sets the size of millions of ligands to the size of thousands of ligands, in a fast and efficient manner without necessarily losing the most promising drug candidates.

Lastly, we also used the MACCS fingerprints to calculate a similarity matrix in between our top 100 hits from the Autodock Vina, 7RDY screen and the 5000 best hits from the NRG Dock screen. From that we were able to determine 8 ligands from the AutoDock Vina screen that were most similar (Tanimoto similarity over 0.5) to the 5000 best hits from the NRG Dock screen, and 11 ligands from the NRG Dock screen that were most similar top 100 hits from the AutoDock Vina screen.

The 8 ligands from the AutoDock Vina Screen were characterized by generally not adhering to Lipinski's rule of five, high molecular weight (mean: 654 g/mole), neutral physiological charge, and in comparison, with the top 10 hits described in Table 2, lower acceptor count (mean: 6) and slightly higher number of rings in the structure (mean: 7).

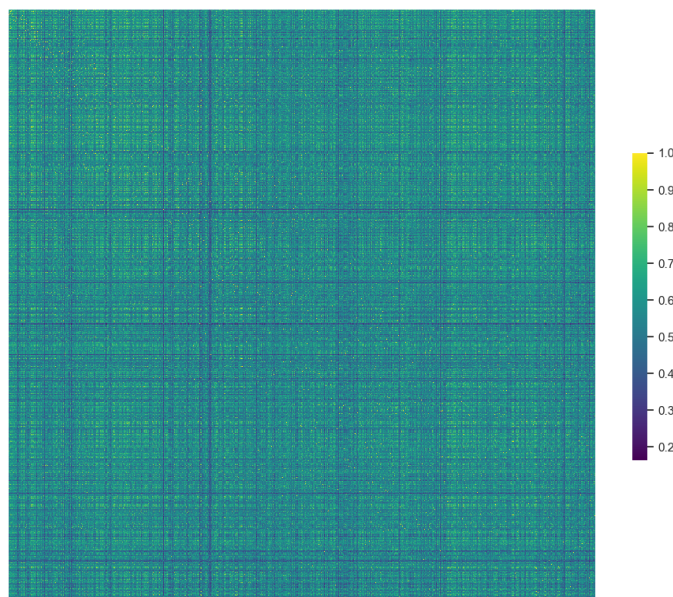


Figure 7. Heatmap of similarity matrix of 1000 ligand subset from the best 5000 hits for each of the binding sites generated by MACCS fingerprints.

We also used the CHEESE Embeddings (*CHEESE*, n.d.) to predict Physiochemical and ADMET properties of the 11 ligands from the NRG Dock screen. Physiochemically the molecules were characterized by mean molecular weight of 434 g/mole, average acceptor count of 7 and neutral formal charge. In terms of their ADMET properties they were characterized by bad Caco2 permeability (mean: -5.09 Log cm/s), good lipophilicity (mean: 2.34 LogD), solubility (mean: -4.98 LogS) and bioavailability (mean: 0.97). They also have a high predicted probability of causing Drug Induced Liver Injury (mean: 0.98), hERG Inhibition (mean: 0.93) and low predicted AMES mutagenicity (0.18). For the full dataset with the predicted AMET properties please see our GitHub. From those predictions we can conclude that even though the molecules are highly similar to approved drugs they don't seem to exhibit the properties necessary for them to have high probability to function as a drug (for example good Caco2 permeability and low toxicity).

4. Discussion

DiffDock, an AI-based molecular docking method, presents promise for virtual screening in drug discovery. However, our study revealed challenges, including errors during screening, especially with larger molecules, highlighting the need for higher computing power. While DiffDock offers unique features like its diffusion generative model and confidence scoring system, discrepancies with AutoDock Vina raise questions about its reliability.

Comparative analysis showed differences in accuracy and efficiency between DiffDock and AutoDock Vina, with no correlation in the results. Methodological improvements of DiffDock are necessary to address errors, such as parameter tuning and integration of experimental validation data. Despite these challenges, optimizing DiffDock could enhance its utility in drug discovery against SARS-CoV-2 and other infectious diseases.

Our study showcases the efficacy of **NRGDock**, a powerful tool for screening vast compound libraries in drug discovery. By employing NRGDock alongside Autodock Vina, we efficiently narrowed down a large pool of compounds while maintaining diversity and potential.

Despite differences in scoring between NRGDock and Autodock Vina, our approach provided valuable insights into ligand diversity and structural characteristics. While discrepancies in rankings were observed, these variations are inherent to different scoring algorithms and do not diminish the effectiveness of NRG Dock in reducing large datasets efficiently.

In essence, NRGDock proves invaluable in accelerating drug discovery efforts. Through its efficient screening capabilities, coupled with comprehensive analyses, we pave the way for the rapid identification and development of novel therapeutics to address critical medical needs.

As was mentioned above, a hard 2-minute cutoff was enforced on all **AutoDock Vina** runs. This was implemented because some runs would take longer than the average 30s (up to 5 minutes which could lead to one molecule being screened for an hour). Had we not implemented it it would have increased the manageable runtime by tens of hours.

Due to the time constraints, there was no other good option. During the first phase of docking (only 7RDY) about 3% (1522) from the DrugBank database had to have been skipped this way. Interestingly, when redocking the top 100 best-scoring molecules another ~30% of the molecules hit the timeout and were skipped. It is not entirely clear why some of the molecules were skipped in the first place. Our theories include that some of the molecules could have been too big and AutoDock Vina was not able to find any suitable conformation in the binding site or it is possible that there were errors in the conversion from .sdf to .pdbqt. More peculiar are the results from the second screening. During the second round of screening the molecules have already passed the first sieve of timeouts which might suggest that the reason why some molecules take much more time to dock lies in the binding pocket. None of these theories were investigated further, however, structural similarity studies and redocking of the skipped molecules (with no timeout) could give us a hint as to why this might have happened.

Another large number of molecules were skipped because AutoDock Vina did not have configurations to dock certain atoms. This resulted in about 5% (2678) loss of molecules from the DrugBank database. From a brief analysis it was due to some more exotic atoms, such as Se, B etc., that were included in the molecules themselves. There is no good way to mitigate this error in the future other than creating custom configurations for AutoDock Vina to recognize these atoms. This error was not propagated to the second screening.

Further, the configuration of the binding pockets for AutoDock Vina was not done with enough rigor. The cube of 20Å around the center of the binding pocket was deemed as large enough with enough excess to accommodate possible binding. It is possible that this was the reason some of the runs took too long as the molecule could not fit into the binding area therefore it was rejected by AutoDock Vina. A possible solution to this could be to create a custom binding area for each molecule which might mitigate the mentioned effect and also in some cases speed up the docking process.

Had we not been constrained by time it would have been possible to do a much more thorough screening of the whole DrugBank database. In our initial analysis of binding sites of the proteins we found 16 different binding sites. Some were of course of low probability, nonetheless, it could have been interesting to analyze them and compare them with the rest. Our initial plans also included the screening of the whole DrugBank database on all of the proteins and all of their binding sites. This data could give us a more thorough understanding of what the certain binding sites prefer and therefore produce a more accurate picture of what drugs could inhibit the Sars-Cov-2 Nsp13 Helicase.

Both our analyses of results from the DrugBank database and the Enamine REAL Diversity Set showed that mostly the molecules did not abide by Lipinski's rule of five. This might suggest that this rule is not as indicative of molecule fitness for drug use as it would seem.

Despite all of these challenges and shortcomings of our research we were successful in finding some interesting molecules that might be useful in drug development for Covid-19. We also showed that there is no correlation between the different docking methods which leads us to conclude that no one docking method should be used for molecule docking without good theoretical background, carefully chosen method and statistically sound results. All of the methods have their place among the others; however, it must be noted that each solves a different problem and research done using these methods must adapt to suit the chosen method.

5. References

- Adedeji, A. O., Marchand, B., te Velthuis, A. J. W., Snijder, E. J., Weiss, S., Eoff, R. L., Singh, K., & Sarafianos, S. G. (2012). Mechanism of Nucleic Acid Unwinding by SARS-CoV Helicase. *PLoS ONE*, 7(5), e36521. <https://doi.org/10.1371/journal.pone.0036521>
- Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., Maruthi, K., Eng, E. T., Vatandaslar, H., Chait, B. T., Kapoor, T. M., Darst, S. A., & Campbell, E. A. (2020). Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, 182(6), 1560-1573.e13. <https://doi.org/10.1016/j.cell.2020.07.033>
- Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*, 61(8). <https://doi.org/10.1021/acs.jcim.1c00203>
- Ivanov, K. A., Thiel, V., Dobbe, J. C., van der Meer, Y., Snijder, E. J., & Ziebuhr, J. (2004). Multiple Enzymatic Activities Associated with Severe Acute Respiratory Syndrome Coronavirus Helicase. *Journal of Virology*, 78(11), 5619–5632. <https://doi.org/10.1128/jvi.78.11.5619-5632.2004>
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z., & Rao, Z. (2019). Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Research*, 47(12), 6538–6550. <https://doi.org/10.1093/nar/gkz409>
- Knox, C., Wilson, M., Klinger, C. M., Franklin, M., Oler, E., Alex, Pon, A., Cox, J. J., Chin, N., Strawbridge, S. A., Marysol García-Patiño, Krüger, R., Aadhavya Sivakumaran, Sanford, S., Doshi, R., Nitya Khetarpal, Omolola Temitope Fatokun, Doucet, D., Zubkowski, A., & Dorsa Yahya Rayat. (2023). DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research*, 52(D1), D1265–D1275. <https://doi.org/10.1093/nar/gkad976>
- Newman, J. A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão-Neto, J., Dunnett, L., Gorriestone, T., Skyner, R., Fearon, D., Schapira, M., von Delft, F., & Gileadi, O. (2021). Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nature Communications*, 12(1), 4848. <https://doi.org/10.1038/s41467-021-25166-6>
- O’Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1). <https://doi.org/10.1186/1758-2946-3-33>
- Romeo, I., Ambrosio, F. A., Costa, G., Corona, A., Alkhatib, M., Salpini, R., Lemme, S., Vergni, D., Svicher, V., Santoro, M. M., Tramontano, E., Ceccherini-Silberstein, F., Artese, A., & Alcaro, S. (2022). Targeting SARS-CoV-2 nsp13 Helicase and Assessment of Druggability Pockets: Identification of Two Potent Inhibitors by a Multi-Site In Silico Drug Repurposing Approach. *Molecules*, 27(21), 7522. <https://doi.org/10.3390/molecules27217522>
- Tanner, J. A., Watt, R. M., Chai, Y.-B., Lu, L.-Y., Lin, M. C., Peiris, J. S. M., Poon, L. L. M., Kung, H.-F., & Huang, J.-D. (2003). The Severe Acute Respiratory Syndrome (SARS) Coronavirus NTPase/Helicase Belongs to a Distinct Class of 5' to 3' Viral Helicases. *Journal of Biological Chemistry*, 278(41), 39578–39582. <https://doi.org/10.1074/jbc.c300328200>
- Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). Diffdock: Diffusion steps, twists, and turns for molecular docking. *arXiv preprint arXiv:2210.01776*.
- NRGDock: An Open-Source High-Throughput Docking Software with Sub-Second Molecule Processing Time. (2023, May 17). Fourwaves. Retrieved January 29, 2024, from <https://event.fourwaves.com/cib2023/abstracts/71a07960-fb5c-4d27-8892-afbdb000334a>
- Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.
- Krivák, R., & Hoksza, D. (2018, August 14). P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1). <https://doi.org/10.1186/s13321-018-0285-8>
- Chen, J., Wang, Q., Malone, B., Llewellyn, E., Pechersky, Y., Maruthi, K., Eng, E. T., Perry, J. K., Campbell, E. A., Shaw, D. E., & Darst, S. A. (2022). Ensemble cryo-EM reveals conformational states of the nsp13 helicase in the SARS-CoV-2 helicase replication–transcription complex. *Nature*, 29(3), 250–260. <https://doi.org/10.1038/s41594-022-00734-6>

- Piplani, S., Singh, P. K., Winkler, D. A., & Petrovsky, N. (2022). Potential COVID-19 Therapies from Computational Repurposing of Drugs and Natural Products against the SARS-CoV-2 Helicase. *International Journal of Molecular Sciences*, 23(14), 7704. <https://doi.org/10.3390/ijms23147704>
- Khanna, V., & Ranganathan, S. (2009, December 3). *Physiochemical property space distribution among human metabolites, drugs and toxins*. PubMed Central (PMC). <https://doi.org/10.1186/1471-2105-10-S15-S10>
- CHEESE. (n.d.). CHEESE. <https://cheese.themama.ai>