Screening for Sars-Cov-2 helicase inhibitors

Anna Audit, Hortense Beaussant, Célia Messaoudi

January 31, 2024

Abstract

After the 2020 COVID-19 pandemic, severe acute respiratory coronavirus 2 (Sars-Cov-2) and proteins associated with it became the main targets for vaccines and, more recently, drugs. One approach aims to prevent the virus from replicating. Therefore, novel drugs directed at the molecules involved in the proliferation of the virus are being investigated. In particular, potential inhibitors targeting helicases such as Sars-Cov-2 helicase (Nsp13). It is impossible to test all possible inhibitors due to both financial and time constraints. Therefore developing new computational method to not only automatically detect potential candidates but also score and rank those candidates has become paramount. In this paper, we built a pipeline to select potential candidates, generate multiple conformations for those candidate, dock them and rank them. Finally, we performed molecular dynamic simulation (MD simulation) on our best performing candidates in order to refine our ranking.

1 Introduction

Helicases are ubiquitous enzymes highly involved in nucleic acid metabolism. They are found in all living organisms, from prokaryotes to eukaryotes, and viruses. Using the energy from nucleotide triphosphate hydrolysis, they are capable of unwinding nucleic acids (DNA and RNA).[11] This results in two single strands of nucleic acid (ssDNA or ssRNA) which can then be used for genome replication. The role of helicases in genome replication, transcription, recombination and repair is essential as impaired helicase function can lead to genetic diseases [4].

Nsp13 is the helicase found in Sars-Cov-2. As stated before, it is necessary for the virus to replicate. Therefore, targeting Nsp13 would prevent the virus from replicating and proliferating [12]. Thus, forestalling the infection. Many studies have been conducted on this protein with different approaches. Here, we aim to develop a basic pipeline for drug screening. By utilizing existing tools and trying to connect them together, we want to build a user-friendly workflow. In the context of Nsp13 inhibitors search, we will use the pipeline thus created to find candidate inhibitors from the European Chemical Biology Database (ECBD).

2 Materials and Methods

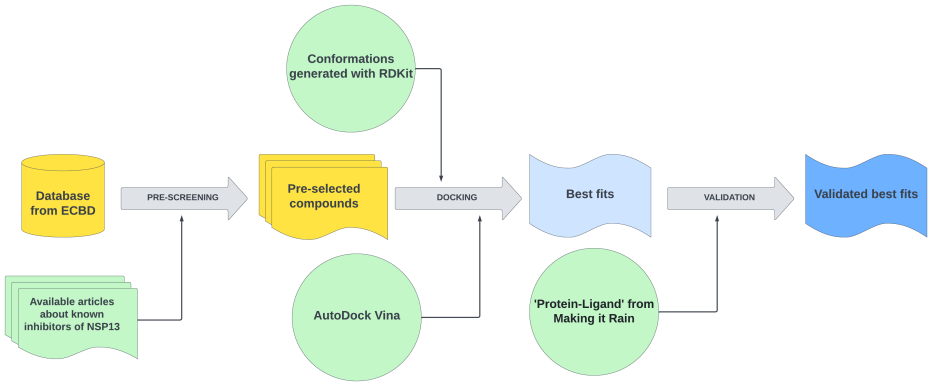


Figure 1: **Inhibitor screening pipeline**. It consists of four main steps. The pre-filtering of our database based on information found in the literature. Then, the generation of multiple conformations for our selected compounds with RDKit, the docking of those multiple conformations with the helicase using Autodock Vina and the subsequent ranking of the conformer based on their affinity score. Finally, MD simulation was performed for the top performing inhibitors.

2.1 Database of compounds

The database used for this project is the pilot library from ECBD. It contains 5015 compounds. As it is a very large number of molecules to test, we decided to do a pre-selection. We only kept molecules that met the following criteria :

- availability in ZINC database,
- logP < 4 and
- $430 \leq$ molecular weight ≤ 470 kDa

The candidates being in the ZINC database is way of ensuring they are available to buy and use in virtual screening campaigns like this one. A low logP value means the inhibiting effect is due to the inhibitor itself and not a "group effect" or "aggregation effect". Molecules with a high logP value are lipophilic and tend to assemble and form an aggregate when highly concentrated [10]. As a result, the effect observed is due to the aggregate blocking the receptor and not a specific action of our molecule in the catalytic site of said receptor. The threshold value of 4 was chosen in agreement with studies done on patented inhibitors of Nsp13 [2]. Since many known inhibitors are small molecules, we chose to focus on candidates with a small molecular weight hence the restriction on this feature [7]. The curated database now contains 276 compounds.

2.2 Generating 3D conformations from SMILES

The following step was generating 3D conformations of our candidates. Using the information contained in the database, we decided it would be a good idea to try and generate the conformations of the ligands ourselves. We used RDKit tools for this step. It allowed us to get conformers using only the SMILES formula provided in the database.

SMILES is a compact way of naming a molecule. It contains information about the atoms and bonds in a molecule and fits in one line. A SMILES is more unique than a usual chemical formula as it is specific to a molecule. RDKit can use this as an input and provide 2D and 3D conformations of the associated chemical compound. In our case, we made a function that does exactly that and saves your conformation as a PDB file. However, RDKit gives random coordinates to the atoms and then optimizes the molecule according to the information contained in the SMILES and its distance optimizing function. Because of this, an infinite number of conformers can exist. To get round this problem, we generated 3 conformers (confo 0-2) for each candidate. This way, we could test

various conformations and see how different the conformations and the results would be. With our 276 candidates, we got 828 conformations in total.

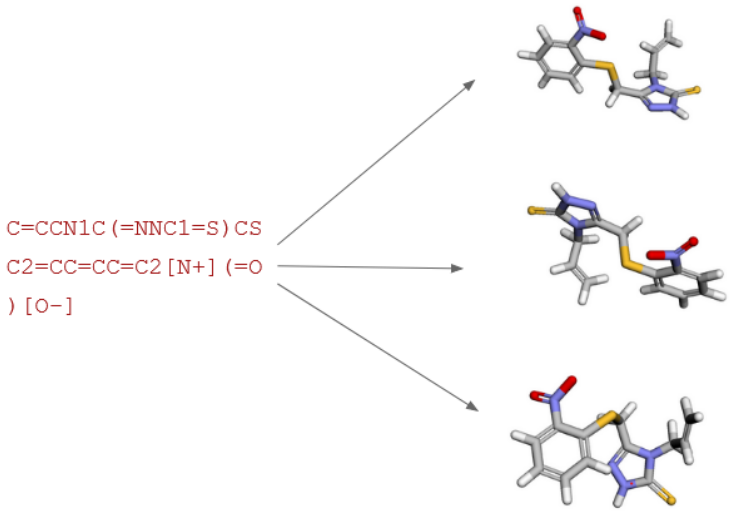


Figure 2: **Example of conformation generated from one SMILES formula**

2.3 Molecular docking

We chose AutoDock Vina [13] as our docking software because it is practical and commonly used in docking studies. Given a protein, its ligand and a binding site, the software is able to generate poses and rank them. AutoDock Vina performs various translations and rotations of the molecule in the binding site to find the pose with the best score. The default Vina affinity score is used here. The best pose is the one with the best affinity score i.e. the most negative one.

Before docking, we did the preparation of the ligands using MGLTools from AutoDock, and its script prepare_ligand4.py. Docking of all 828 conformations was done on the 7nio conformation of Nsp13 and the 10 best scoring candidates were kept. We let Vina determine automatically the binding sites after the parametrization of the grid of the protein by MGLTools. The following parameters were used for the docking :

$center_x = -20.150$
$center_y = 11.494$
$center_z = -7.319$
$size_x = 40$
$size_y = 40$
$size_z = 40$
$energy_range = 4$
$exhaustiveness = 8$

We obtain this kind of results, where we have 9 different poses for each docking :

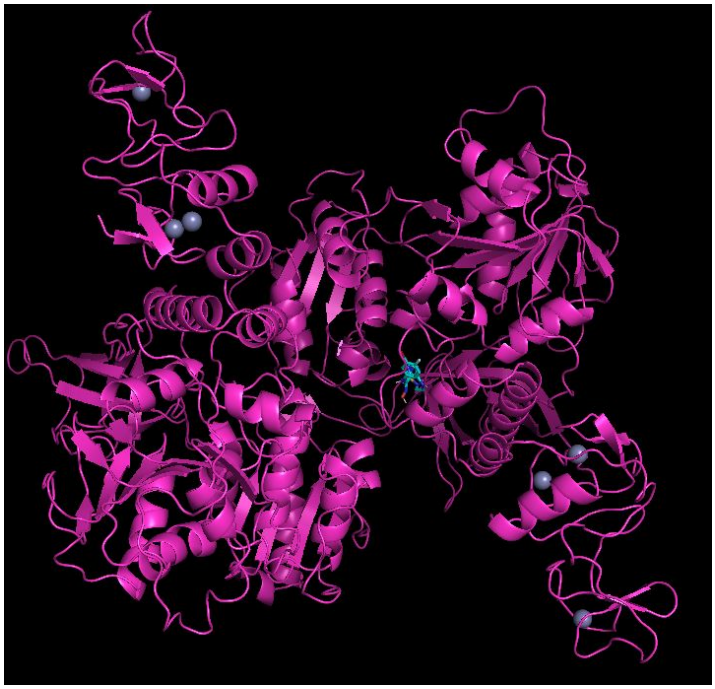


Figure 3: **Example of results from docking, the protein is in pink and the ligand, almost in the center in blue**

2.4 Molecular Dynamics simulation

Molecular Dynamics (MD) simulations were run for the top candidates found after docking to get a dynamic representation of the system over time. MD simulations take into account the flexibility of the molecules and conformational changes whereas classical docking considers the molecules to be more rigid. This step allowed us to refine our previous ranking.

We used Making-it-Rain 'Protein-Ligand' notebook [3]. It contains a step-by-step guide to MD simulations and is user-friendly. The pipeline uses OpenMM engine and AMBER force field and only needs PDB files of the protein and the ligand as inputs. To generate the topologies, it uses ff19SB force field for the protein (Nsp13) and GAFF2 for the ligand. We simulated 10 ns of interaction for each compound using the default parameters of the notebook. The analysis functions provided RMSD (Root Mean Square Deviation) , MM-GBSA (Molecular Mechanics Generalized Born Surface Area) and MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) scoring among others. As we focused on small molecules, RMSD and MM-PBSA scores were kept and used for the ranking. Also, this step was only done on the top 5 candidates out of the 10 we got from the docking because of computational resources and time constraints. The notebooks are made to be used in Google Colab and the free version has time limits for long simulations. The whole pipeline was tested on a known inhibitor of Nsp13 - SSYA10-001 - to make sure the values and parameters used were leading us the right way.

The RMSD, MM-PBSA and MM-GBSA analysis are complementary and give us different types of information [6]. The RMSD allows us to monitor the dynamic behavior and stability of the protein-ligand complex over time, while MM-PBSA and MM-GBSA are utilized to estimate the binding free energy and understand the thermodynamics of biomolecular interactions.

The RMSD computes how much a structure deviates or "moves" from a reference position over the course of a simulation. This gives an estimate of the stability of the system. A stable complex will typically have a relatively constant RMSD while large and sudden variation of the RMSD indicates instability. The formula for the RMSD is : $RMSD = \sqrt{\frac{1}{N}\sum_N (r_i - r_{i,ref})^2}$. With r_i the position of the atom i.

The MM-PBSA is computational method used to estimate the free-binding energy of the protein-ligand

interaction. The MM-PBSA has three main components:

- Molecular Mechanics (MM) Energy, calculates the gas-phase interaction energy between the protein and ligand. This includes terms such as van der Waals, electrostatic, and bonded interactions.
- Poisson-Boltzmann (PB) Solvation Energy accounts for the interactions between the biomolecular complex and the solvent. This estimates the solvation energy.
- Solvent-Accessible Surface Area (SASA), the nonpolar solvation energy is estimated based on the solvent-accessible surface area (SASA) of the complex.

The free binging energy ΔG_{bind} is then computed using the formula : $\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{protein} + \Delta G_{ligand})$.
The MM-GBSA is very similar to the MM-PBSA method, expects it uses a different solvation model (GB instead of Poisson-Boltzmann) to estimate the solvation effects.

3 Results

3.1 Molecular docking

Since the 7nio conformation of Nsp13 is a homodimer, two docking sites are visible. They are symmetrical i.e. they are identical but located on one half of the dimer each.

Figure 4: Docking site 1

Figure 5: Docking site 2

3.2 Best scoring compounds after docking

As seen in Table 1, our 10 best scoring compounds all have an affinity score around or lower than -10 kcal/mole. If we look at the 15 best candidates, we find some compounds 3 times, one for each conformation we generated. It is the case for our best scoring ligand (ZINC000049006633) in Table 1. To test a bigger variety of ligands, we kept the best scoring conformation for each ligand in the top 15 and got our top 10 (Table 1). However, the scores are very close so we can't conclude on some candidates being way better than others. Also, we expect a similar but not identical ranking after MD simulations. For molecules such as ZINC000049006633, we expect good results in MD simulations too since the affinity seems to be very good.

3.3 Best scoring compounds after MD simulations

The ranking obtained (Table 2) is a little different from the first one (Table 1). ZINC000043208642 seems to be a better option here than before (top 5 to top 1) whereas ZINC000049006633 (top 1

Formula	ZINC ID	Affinity score (kcal/mole)	Place
C24H17N7O3 confo 0	ZINC000049006633	-11.1	1
C21H22FN3O5S confo 2	ZINC000064889585	-10.9	2
C21H22ClN3O5S confo 2	ZINC000064889599	-10.9	2
C24H33FO6 confo 1	ZINC000064889599	-10.8	4
C24H31FO6 confo 0	ZINC000004097305	-10.8	4
C24H32ClFO5 confo 0	ZINC000004213474	-10.7	6
C25H25N5O4 confo 0	ZINC000011677837	-10.7	6
C25H30N6O3 confo 1	ZINC000059258964	-10.7	6
C27H27N5O3 confo 2	ZINC000033173150	-10.6	9
C22H20F5N3O3 confo 2	ZINC000043208642	-10.6	9

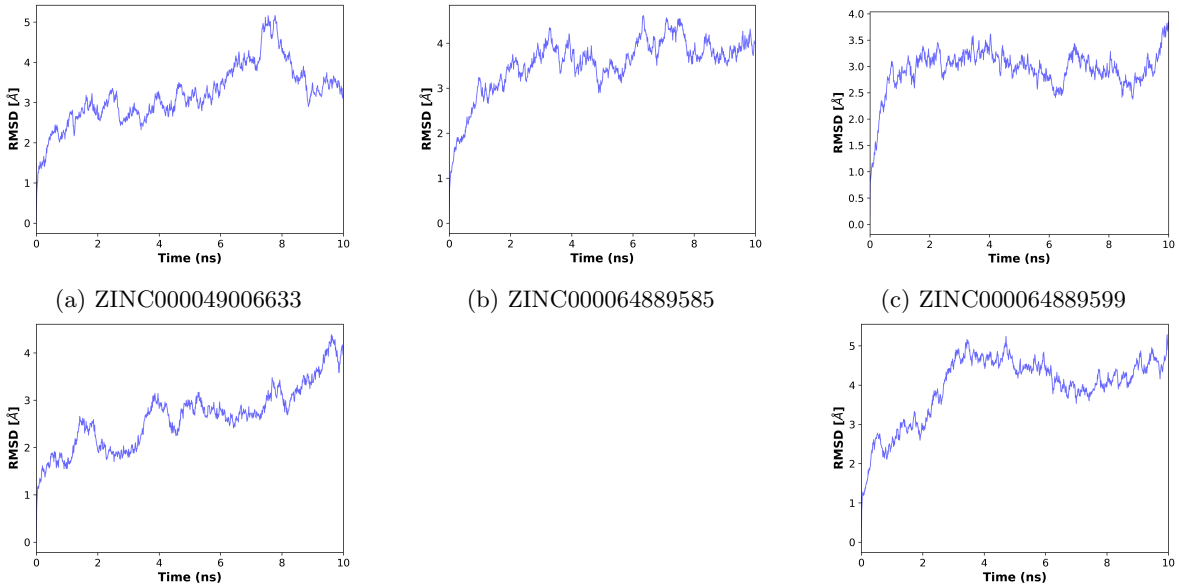
Table 1: Top 10 candidates after docking

before) is in the lower part of the ranking. This could be explained by the conformational changes the different molecules can undergo while interacting with Nsp13. Some molecules are more flexible than others. Looking at the RMSD scores, most are around 3 Åwhich is considered a good value in similar drug screening studies. We should also note that the variability in the MM-PBSA scores is larger than the one for MM-GBSA scores. A possible reason for that is the solvation parameters used during MD simulations. In fact, MM-PBSA scoring is more sensitive to solvation parameters than MM-GBSA scoring. As a result, MM-PBSA is more accurate when MM-GBSA is faster but makes some simplifications. Another remark is the positivity of most MM-PBSA scores which can be explained by the parameters used in the MD simulations or the just the instability of the complex formed. Here, both scores were computed to see the differences. Our interpretations and ranking are based on the MM-PBSA values to give more importance to the accuracy of the calculation.

Formula	ZINC ID	MM-GBSA (kcal/mole)	MM-PBSA (kcal/mole)	RMSD (Å)	Place
C24H17N7O3 confo 0	ZINC000049006633	-6.9806	4.6739	2.78 ± 0.69	3
C21H22FN3O5S confo 2	ZINC000064889585	-3.1628	4.2446	3.53 ± 0.64	2
C21H22ClN3O5S confo 2	ZINC000064889599	-8.2222	4.7031	2.99 ± 0.62	4
C24H33FO6 confo 1	ZINC000004097308	-8.7427	5.0721	2.64 ± 0.67	5
C24H31FO6 confo 0	ZINC000043208642	-9.7767	-1.8422	3.94 ± 0.85	1

Table 2: Top 5 candidates after MD simulations

As we can see on figure 6, not all structures seem to be very stable. The RMSD scores keep increasing with time, except for ZINC000064889599 and ZINC000064889585. For these two molecules, the RMSD values are little more stable and their standard deviation is also a little lower than for others (Table 1). This can be due to the length of the simulation (only 10 ns), the values may converge towards a more stable value if we run the simulation for longer.



(a) ZINC000049006633 (b) ZINC000064889585 (c) ZINC000064889599

(d) ZINC000004097308 (e) ZINC000043208642

Figure 6: RMSD scores during MD simulation

3.4 Biological implications

After doing some investigation on some of the compounds, we found that some of them were already known as inhibitors for other processes and pathways. For example, ZINC000043208642 is used for its anti-inflammatory role as a Notch pathway inhibitor [8, 9]. Our results show it could be repurposed as a Nsp13 inhibitor, after experimental validation. Similar conclusions can be drawn for each ligand in our top 5 (or top 10) because they are all available in the ZINC database, therefore already used for other issues.

3.5 Comparison with other teams

Our top ligands are different from the ones found by other teams, even though the Heidelberg team used a mix of ECBD and ZINC databases too. This can be due to the parameters used in the docking step. Also, our affinity scores are close but a little higher than theirs. Same observation for affinity scores from the Warsaw 1 team, they have higher values. The use of various databases and docking softwares (some teams used DiffDock [5]) may explain the differences in our results. Some of us also used different conformations of Nsp13 such as 7nn0 and 7nio.

4 Discussion

The pipeline we used allowed us to screen a large library of compounds and choose the ones that fit best with Nsp13. However, some steps could still be improved. For example, trying DiffDock [5] instead of AutoDock Vina to compare the docking results. Especially, to see if the same poses are obtained and if the rankings are similar.

About MD simulations, the recent updates of Google Colab make it almost impossible to run the whole notebook without the Pro version. Therefore, using other tools like GROMACS[1], NAMD, AMBER and OpenMM [3] can be a good alternative. The time and resources needed for 10 ns of simulation are still very large, whatever the tool is. Here, we used Making-it-Rain because using the cloud resources of Google Colab allows us to run a whole simulation, from importing the files to running the analysis of free binding energy, in around 24 hours when it takes GROMACS a day and a half. Deep learning methods such as DeePMD-kit [14] can also be used to study biomolecular interactions instead of classical methods. Nevertheless, these types of algorithms are still in early development stages and might not be as accurate as the ones cited previously yet.

Another area to improve would be the parameters used at various steps, notably docking and MD simulations. Using a smaller simulation box for docking or changing the solvation parameters for equilibrium before MD simulation may give us more precise and accurate results. Also, we only simulated 10 ns. Simulations of 150 ns or more would allow us to assess the stability of the interaction better, especially when interpreting the RMSD scores measured throughout the interaction time. Finally, using the same database would make the comparison of our results with other teams easier.

References

[1] Mark James Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (Sept. 2015), pp. 19–25. ISSN: 2352-7110. DOI: 10.1016/j.softx.2015.06.001. URL: https://www.sciencedirect.com/science/article/pii/S2352711015000059 (visited on 01/31/2024).

[2] Yazdi Ak et al. "Kinetic Characterization of SARS-CoV-2 nsp13 ATPase Activity and Discovery of Small-Molecule Inhibitors". en. In: *ACS infectious diseases* 8.8 (Dec. 2022). ISSN: 2373-8227. DOI: 10.1021/acsinfecdis.2c00165. URL: https://pubmed.ncbi.nlm.nih.gov/35822715/ (visited on 11/21/2023).

[3] Pablo R. Arantes et al. "Making it Rain: Cloud-Based Molecular Simulations for Everyone". In: *Journal of Chemical Information and Modeling* 61.10 (Oct. 2021), pp. 4852–4856. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.1c00998. URL: https://doi.org/10.1021/acs.jcim.1c00998 (visited on 01/01/2024).

[4] A. J. van Brabant, R. Stan, and N. A. Ellis. "DNA helicases, genomic instability, and human genetic disease". eng. In: *Annual Review of Genomics and Human Genetics* 1 (2000), pp. 409–459. ISSN: 1527-8204. DOI: 10.1146/annurev.genom.1.1.409.

[5] Gabriele Corso et al. "DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking". en. In: Sept. 2022. URL: https://openreview.net/forum?id=kKF8_K-mBbS (visited on 01/31/2024).

[6] Samuel Genheden and Ulf Ryde. "The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities". In: *Expert Opinion on Drug Discovery* 10.5 (May 2015), pp. 449–461. ISSN: 1746-0441. DOI: 10.1517/17460441.2015.1032936. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487606/ (visited on 01/31/2024).

[7] Zeng J et al. "Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of nsp13 helicase". en. In: *The Biochemical journal* 478.13 (July 2021). ISSN: 1470-8728. DOI: 10.1042/BCJ20210201. URL: https://pubmed.ncbi.nlm.nih.gov/34198322/ (visited on 11/21/2023).

[8] Rajesh Kumar, Lucienne Juillerat-Jeanneret, and Dela Golshayan. "Notch Antagonists: Potential Modulators of Cancer and Inflammatory Diseases". In: *Journal of Medicinal Chemistry* 59.17 (Sept. 2016), pp. 7719–7737. ISSN: 0022-2623. DOI: 10.1021/acs.jmedchem.5b01516. URL: https://doi.org/10.1021/acs.jmedchem.5b01516 (visited on 01/29/2024).

[9] Leopoldo Luistro et al. "Preclinical profile of a potent gamma-secretase inhibitor targeting notch signaling with in vivo efficacy and pharmacodynamic properties". eng. In: *Cancer Research* 69.19 (Oct. 2009), pp. 7672–7680. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-09-1843.

[10] Kenny Pw, Montanari Ca, and Prokopczyk Im. "ClogP(alk): a method for predicting alkane/water partition coefficient". en. In: *Journal of computer-aided molecular design* 27.5 (May 2013). ISSN: 1573-4951. DOI: 10.1007/s10822-013-9655-5. URL: https://pubmed.ncbi.nlm.nih.gov/23737238/ (visited on 11/21/2023).

[11] Anna Marie Pyle. "Translocation and unwinding mechanisms of RNA and DNA helicases". eng. In: *Annual Review of Biophysics* 37 (2008), pp. 317–336. ISSN: 1936-122X. DOI: 10.1146/annurev.biophys.37.032807.125908.

[12] Austin N. Spratt et al. "Coronavirus helicases: attractive and unique targets of antiviral drug-development and therapeutic patents". In: *Expert Opinion on Therapeutic Patents* 31.4 (Apr. 2021), pp. 339–350. ISSN: 1354-3776. DOI: 10.1080/13543776.2021.1884224. URL: https://doi.org/10.1080/13543776.2021.1884224 (visited on 01/25/2024).

[13] Oleg Trott and Arthur J. Olson. "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 (Jan. 2010), pp. 455–461. ISSN: 0192-8651. DOI: 10.1002/jcc.21334. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041641/ (visited on 01/31/2024).

[14] Han Wang et al. "DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics". In: *Computer Physics Communications* 228 (July 2018), pp. 178–184. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2018.03.016. URL: https://www.sciencedirect.com/science/article/pii/S0010465518300882 (visited on 01/30/2024).