MeetEU Project - Team Heidelberg - Team 1 – Identification of novel Sars-CoV-2 NSP13 helicase inhibitors

Anastasiya Vladimirova, Darius Szablowski, Duc Thien Bui, Wangjun Hu

February 2023

1 Abstract

As new variants of the Sars-CoV-2 virus have emerged in recent, more resources have been shifted towards the goal of identifying potential inhibitors for this infection. As experiments in the laboratory take too much time, are susceptible to human error and can only cover so many molecules due to monetary issues it has become a standard to firstly identify potential inhibitors using a computational approach [Luttens et al., 2022].

We succeeded in building a pipeline which is capable of selecting potential inhibitors with preselection using requirements such as their toxicity. In the future other preconditions just like the one of it having a low cost could also be implemented further reducing the computing time.

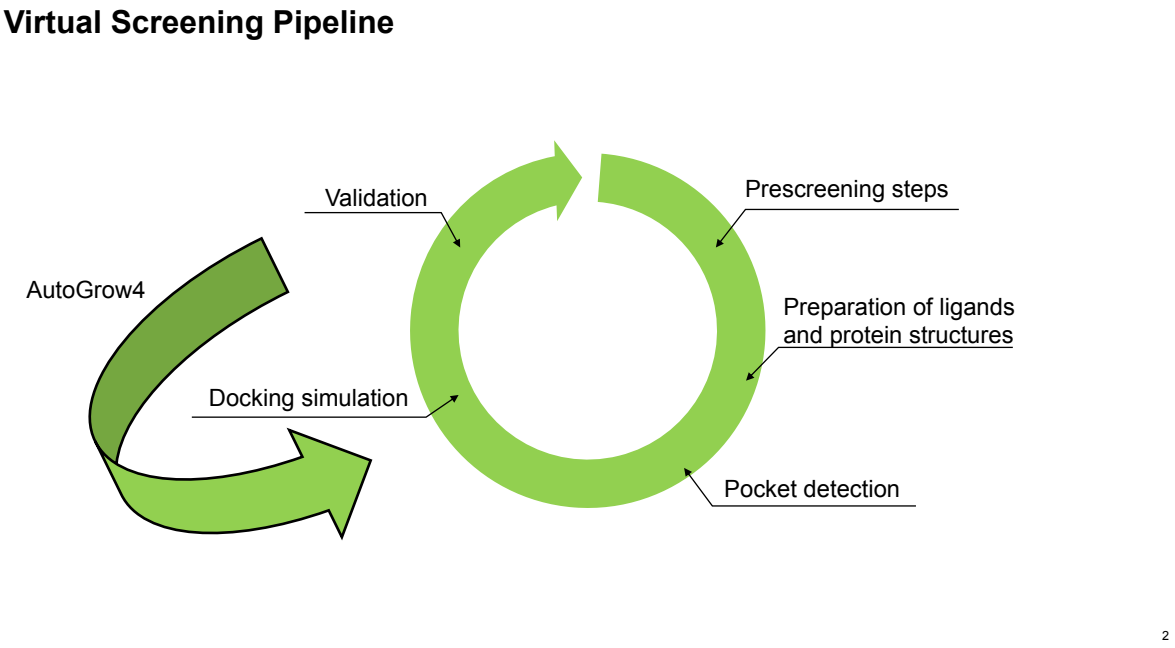


Figure 1: **Virtual Screening Pipeline.** It consists of the following steps: Prescreening steps, e.g. Toxicology Check, Preparation of ligands and protein structures with *PyMOL* by removing water molecules and converting to the correct file format, Pocket detection via *Fpocket*, Docking simulation with *AutoDock Vina* using the *Vinardo* scoring function, using the results from that to assimilate new ligands using *AutoGrow 4* to rerun the Docking simulation and lastly Validation using Molecular Dynamics Simulations with *GROMACS*.

2 Introduction

For the implementation of our pipeline there were several minor goals which were given as a guideline by the Meet-EU course administration:

1. Characterizing druggable binding sites of the enzyme
2. Investigating their interactions with known inhibitors
3. Developing predictive models for virtual screening campaigns
4. Performing virtual screening simulations on the identified binding po-ckets

The final judging criteria was the virtual screening of the Pilot library of the European Chemical Biology Database (ECBD) which contains 5001 molecules.

2.1 Characterizing druggable binding sites of the enzyme

To perform molecular docking simulations, one needs to first define a box in which the molecular docking takes place [Eberhardt et al., 2021]. For a receptor ligand docking it is the site of the binding sites. There are two different binding sites on the NSP13 protein, namely the ATP binding site and the DNA-/RNA-binding site [White et al., 2020].
Since more recent papers concentrated on the ATP-binding site we decided to do so as well to reduce the computing time or computing power needed for the analysis [White et al., 2020].
After selection of the druggable site further preparation needs to be done, e.g. checking for the quality of the structure and converting the file into the right format for further analysis.

2.2 Investigating their interactions with known inhibitors

The analysis of known inhibitors is a good way to perform a test on any given pipeline and to further improve it by integrating the most important residues. This is also a simple way to reduce the needed computing power or computing time since the box which the simulation is being performed in can be minimized [Pushpakom et al., 2018].

As we will only concentrate on the ATP binding site of the NSP13-protein the interactions of the ligands will only have to be evaluated with one binding site.

2.3 Molecular Docking

For the molecular docking of the ligands several circumstances have to be considered, e.g. the rotation and translation of each ligand, meaning that for each ligand, several simulations have to be performed to investigate the best possible rotation and translation of each ligand before then being able to compare the ligands' results with one another. The value, which measures how high the affinity between ligands and the receptor is, is called docking score. The docking pose with the highest docking score will be the best docking pose and similarly, the ligand with the best docking score will therefore be estimated to be the best ligand overall [Al-Karmalawy et al., 2021].

There are several approaches to identifying the best complex but as the checking for every possible pose for every ligand would need an extensive amount of time and computing power there are several search algorithms within tools which can be used to speed up the search and make it more efficient. These tools use so called scoring functions which are capable of discriminating between correct and incorrect results to facilitate the search for the poses. As simplifications and assumptions are being made the scoring functions don't necessarily calculate the exact binding affinity but rather rank the poses according to binding affinity [Guedes et al., 2018].

2.3.1 Scoring functions

There are a multitude of scoring functions which can be used for the calculation of the ranking. In brief, they can all be categorized into four different groups:

1. **Force field based:** A forcefield is used to evaluate the binding affinity between the ligand and the receptor, e.g. *CHARMM*, *AMBER* used *AutoDock* or *CGenFF* by *EnzyDock*
2. **Empirical:** Various important energetic factors are being weighted and taken into consideration, e.g. ChemScore or X-Score
3. **Knowledge based:** Using the frequency of a certain interaction is estimated to be a unit of measure for the importance of said interaction, e.g. DrugScore or PMF
4. **Machine-learning based:** Several docking-based parameters, structural interaction fingerprints and atom pairs are used to develop AI-based scoring functions which are weighing the different influences according to their training data set.

[Guedes et al., 2018]

2.3.2 Developing predictive models for virtual screening campaigns

Using the findings of the interactions of known inhibitors with binding pockets, it is possible to train predictive models which are then capable of selecting possible inhibitors. One example would be to minimize or enlarge the box in which the simulation takes place depending on prior results. Just like that with tweaking of certain parameters the specificity of the model can be maximized by training it with or testing it on other data sets [Joshi et al., 2020].

2.3.3 Performing virtual screening simulations on the identified binding pockets

In general, virtual screening aims to discriminate between active and inactive compounds and thus select the best possible inhibitors for a given protein [Joshi et al., 2020].

2.4 Generating New Molecules

An important branch of drug discovery is molecule generation. Computational drug design approaches owe their success to the development of machine learning methods. They are varied in their nature and are based on different types of machine learning, but the most successful approaches are based on deep learning - several processing layers that are able to extract complicated features. Compared to other algorithms, deep learning has the ability to capture non-linear dependencies. The field can be separated into three categories: de-Novo drug design/optimization, synthesis prediction, and structure-based modeling. Depending on the molecular representation, they can also be classified as atom-based, reaction-based, and fragment-based. There are an overwhelming amount of algorithms available and it can be challenging to choose an appropriate one for the task. However, with the help of reviews like [Meyers et al., 2021] and [Jiménez-Luna et al., 2021], the search field can be significantly reduced.

Our plan is to utilize deep learning in optimizing the best-performing molecules and potentially discovering a better fit for the pocket.

Autogrow 4, presented by [Spiegel and Durrant, 2020], is the algorithm we selected for the task. It's a genetic algorithm, that is reaction-based: it utilizes the library of reactions to mutate molecules and create a new generation.

Autogrow 4 is not limited to the lead optimization, it can also perform de novo molecule generation and the functionality depends on the input of the algorithm, the broader the original generation, the more algorithm leans towards de-Novo generation.

The algorithm works as follows: from the initial generation (or population) several molecules are selected and the new generation is created which is then docked and the results of it are used to assess the best-scoring molecules, which are then fed into the next generation. The new population is created via elitism, mutation, and crossover. The elitism function is the selection of the top-scoring docked molecules of the previous generation. The mutation is used to generate compounds with in silico chemical reactions. It could potentially be problematic, given that Autogrow 4 does not consider other reactive groups in the molecule when performing a mutation on a specific reaction group. However, it can be overlooked since it creates a necessary diversity. The last operator - crossover - combines similar molecules and randomly keeps/remove their respective moieties to the common substructure.

Molecules are filtered by their physical and chemical properties before they are fed into the docking and then the entire process repeats.

The AutoGrow 4 was not only selected due to its novel approach to molecule generation and good performance on the benchmarking sets but also due to its user-friendliness. Autogrow is fast, modular, and well-documented. It is implemented in python, allows to run it in docker, and is easily

extendable. For example, the molecule conversion from pdb to pdbqt can be done with OBabel [O'Boyle et al., 2011], as well as MGLTools [S, 2010]. The docking per default is performed with Autodock Vina, but for faster runs, one can specify the use of QuickVina [Handoko et al., 2012]. The docking algorithm can also be completely replaced by another algorithm of the user's choosing.

3 Material and Methods

3.1 Structure and Binding Pocket Detection

The 6ZSL[1] structure from the RCSB PDB database was chosen for the detection of binding pockets. The structure was first described in [Newman et al., 2021].

Based on this structure binding pockets were predicted using the Fpocket [Guilloux et al., 2009] and Prankweb[2] [Jendele et al., 2019] tools. Because of the limited scope of this project only one binding pocket was chosen for further consideration.

3.2 Test Dataset from ZINC15

All drug-like molecules with the following parameters were downloaded from the ZINC15-Database [Sterling and Irwin, 2015]. The number of drug-like molecules contained in the ZINC15-Database is 981,247,974. We downloaded a subset of these. Substances downloaded were only the ones having a 3D-representation, reactivity: "Standard", a purchasability status of: "Wait OK", representation pH(s) in the Mid(M) range and charges of -2, -1, +1 and +2. The number of initially downloaded molecules was 95,438,992.

3.3 Toxicology prediction using *eToxPred*

The machine learning based tool *eToxPred* [Pu et al., 2019] was used for toxicology prediction of the given molecules. Based on the information given by the paper the threshold was set to 0.58 for a substance to be considered toxic.

3.4 Docking

For docking 5000 molecules out of the molecules that were predicted to be non-toxic were randomly chosen for docking. This was done because a comprehensive docking would have not been able in the given time. These molecules were randomly chosen from multiple of the downloaded tranches in order to ensure that the checked molecules contained a diverse set of substances and cover as much of the theoretical search space with the limited number of ligands that could be tested. Docking was performed using *AutoDock Vina* [Forli et al., 2016] using the *Vinardo* scoring function [Quiroga and Villarreal, 2016]. *Vinardo* was chosen because it outperformed classical *Vina* scoring on multiple datasets at identifying the top poses of potential ligands. The exhaustiveness was set to 32.

These were the binding site parameters used for docking:

```
center_x = -14.2861
center_y = 11.3817
center_z = -75.3838
size_x = 30.0
```

[1] http://dx.doi.org/10.2210/pdb6zsl/pdb
[2] https://prankweb.cz/

```
size_y = 30.0
size_z = 30.0
```

3.5 PDB Preparation

The crystallized protein structures of NSP13 were downloaded as previously described (see Section 3.1) at RCSB [3]. Missing residues and atoms were added using the Python package *'pdbfixer'* [Eastman et al., 2013]. Additionally, water and phosphate were removed from the structures. Hydrogen atoms were not added as *GROMACS* can add them too. The resulting '.pdb' files were consecutively processed by *VEGA* [Pedretti et al., 2002] to ensure correct PDB format.

3.6 Determining Top 10 and Ligand Preparation

A dictionary of small molecule name and binding energy was given by the *Autodock Vina* package using the *Vinardo* scoring function. The values were then sorted by the strongest binding affinity (the more negative the value, the stronger the binding affinity) and the top 10 molecules with the lowest binding energy were extracted for MD simulation.

The resulting docking poses from the different small molecules were given in a '.pdbqt' format. In this '.pdbqt' file different binding poses are listed. For MD simulation the best model, thus 'Model 1' was chosen. Due to unknown reasons some carbon atoms were not recognized and were marked as G/Du dummy atoms. Hence, further processing was made more difficult. 'Model 1' of the top 10 molecules were visualized using *Avogadro* [Hanwell et al., 2012] and the corresponding native molecules or not docked molecules were downloaded from the Zinc15 Database. By one versus one comparison the missing atoms, bonds and hydrogens were added using *Avogadro*. The prepared ligands were then exported as '.pdb' and '.mol2' files for MD simulation.

3.7 Molecular Dynamics Simulation using *GROMACS*

GROMACS is a tool to conduct molecular dynamics simulation on biological macromolecules such as nucleic acids and proteins [Bekker et al., 1993]. Furthermore, *GROMACS* can be used to study protein-ligand interactions [**?**]. The broad application scenarios of *GROMACS* are well documented in tutorials by Dr. Justin Lemkul([Lemkul, 2019]). In this project, the top 10 protein-ligand pairs from the docking were analysed according to the 'Protein-Ligand Tutorial' [4]. The .pdb and ligands were processed as described in section X and Y. Recent published papers conducting MD simulation on protein-ligands were using *AMBER* forcefields ([Raubenolt et al., 2022, Chen et al., 2022]), hence the *AMBER99SB* [Hornak et al., 2006] force field was selected. The ligand forcefield was generated using the Python package *'acpype'* ([da Silva and Vranken, 2012]), which generates ligand AMBER forcefields. After preparing the forcefields, the standard workflows were conducted including solvation and neutralization of the system, a first energy minimization, equilibrium phase and production MD. We've been using the standard parameter files with default settings suggested by the tutorial.

3.8 Visualization and Analysis of MD simulation results

After adjusting our trajectories according to the tutorial, the protein-ligand simulation can be analyzed. The protein-ligand interaction energy was calculated, resulting in the computation of short-range coulomb and long-range Lennard-Jones energies for each protein-ligand pair. The total interaction energy summing up both terms was computed and used for the second stage of ranking. Additionally, the RMSD between the protein and ligand was computed. The workflows were derived from the protein-ligand tutorial by Dr. Justin Lemkul([Lemkul, 2019]). Furthermore, the 10 ns simulations were visualized using the program *VMD* [Humphrey et al., 1996].

[3] `https://www.rcsb.org`

[4] `http://www.mdtutorials.com/gmx/complex/index.html`

3.9 Second Docking Step

Another Docking round again using *AutoDock Vina* with the *Vinardo* scoring function was performed. As per the task of this project a subset of the European Chemical Biology Database (ECBD)[5], specifcally the subset: `https://ecbd.eu/compound/#lib{value='4'}`.

3.10 Generating New Molecules

The same protein structure and binding pocket as described above is used for the generation of the new molecules. Autogrow 4 performs docking with standard AutoDock Vina and format conversion with MGLTools. From the 5000 thousand docked molecules the top 500 are selected to perform generation of new molecules.

The parameters for the AutoGrow 4 can be found in the JSON file on gitlab. There are 15 generations created by the program.

4 Results

4.1 Binding Pocket identification

4.1.1 *Prankweb*

Prankweb predicted binding pockets					
name	rank	score	center_x	center_y	center_z
pocket1	1	13.01	-14.2861	11.3817	-75.3838
pocket2	2	7.48	-19.4416	34.3608	-22.3538
pocket3	3	5.80	-29.1017	16.8762	-22.2311
pocket4	4	5.24	-22.4480	28.4716	-78.3150
pocket5	5	1.07	4.3560	-7.2945	-48.6003

Table 1: **The name, rank, score, and x, y and z coordinates of the predicted binding pockets.**

4.1.2 *Fpocket*

The three predicted binding pockets with a drugability score bigger than 0.5 were:

$$
\begin{aligned}
Pocket2: &\quad x = -12.31295774647887, &\quad y = 13.357760563380282, &\quad z = -28.824183098591543\\
Pocket3: &\quad x = -15.068947368421053, &\quad y = 13.547719298245612, &\quad z = -73.44014035087719\\
Pocket50: &\quad x = -37.6075, &\quad y = 9.999269230769233, &\quad z = -29.833961538461544
\end{aligned}
$$

4.2 Toxicology prediction using *eToxPred*

Due to several problems and the limited scope only a subset of the previously mentioned molecules from the *ZINC15*-database could be run through the toxicology prediction using *eToxPred*. The total number of molecules we were able to evaluate was 19,666,112. Of these 19,071,206 were predicted non-toxic. This means approx. 96.97% of the molecules were deemed non-toxic and 3.03% were predicted toxic.

[5] `https://ecbd.eu`

4.3 General Docking Results ZINC15 dataset subset

The resulting binding energy of our 5000 starting molecules were loaded visualized in a histogram using R (see Figure 2). Most of the molecules have binding affinities between – 2.5 and -10 $[\frac{kJ}{mol}]$. A small fraction of molecules has binding affinities below – 10 which include our top 10 candidates (see Table 2). Unfortunately, an error occurred during computation, thus we proceeded with the top 9 candidates.

Table 2: **Ranking of *Autodock Vina* poses and reranking based on MD simulation results.** The ranks differ between the binding affinity sorting and the reranking after MD simulation evaluation.

Autodock Vina Top 9		
Ranking	Zinc ID	Ranking New
1	ZINC000514436632	3
2	ZINC000104309836	4
3	ZINC000620748806	2
4	ZINC000019015192	6
5	ZINC000104277568	8
6	ZINC001164872284	5
7	ZINC001164979321	1
8	ZINC000101042701	9
9	ZINC000095523345	7

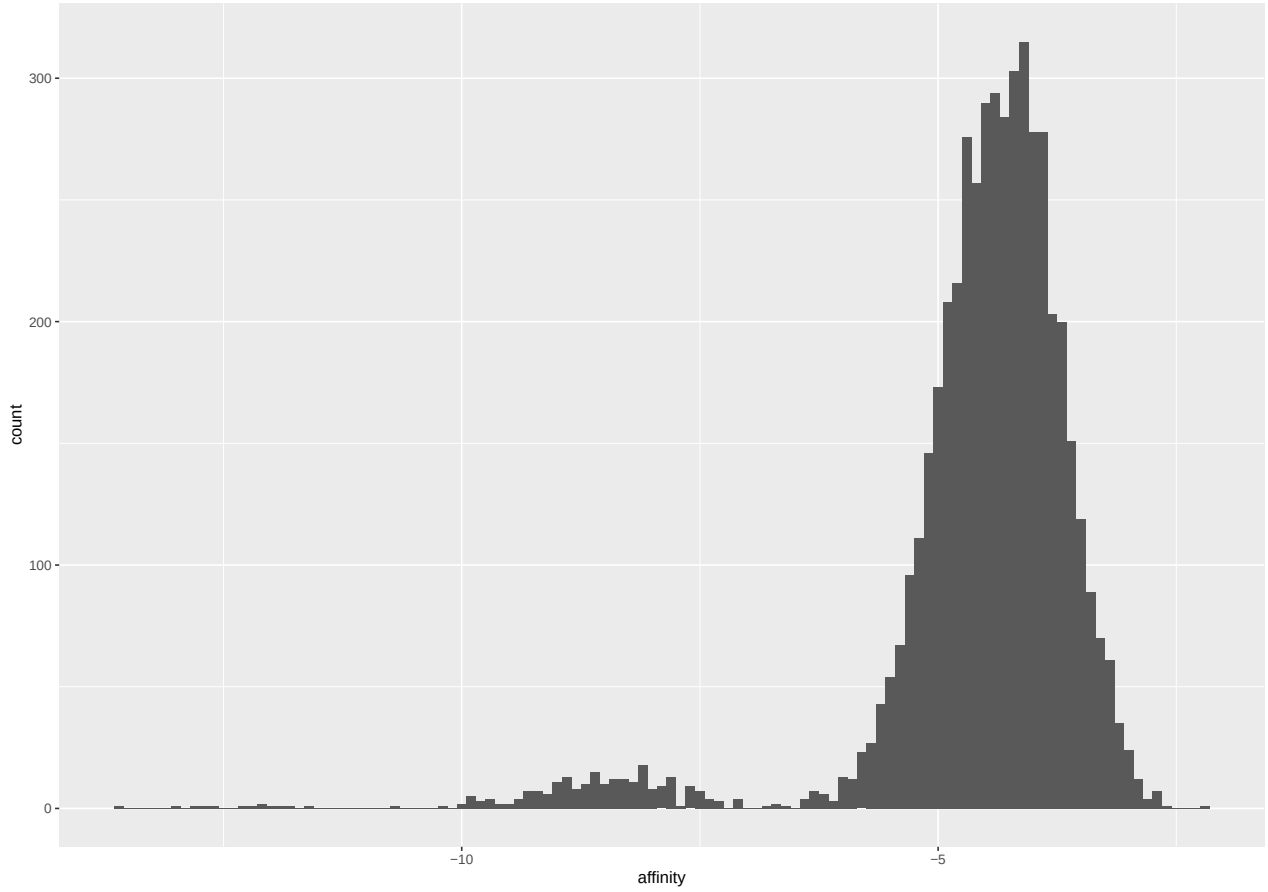


Figure 2: **A histogram of the frequency-counts of the determined binding affinities in $[\frac{kJ}{mol}]$.**

4.4 Reranking our top 10 molecules based on the total protein-ligand interaction based on MD simulation results

The total protein-ligand interaction energy consisting of the Coulomb and Lennard Jones energy was computed and is shown in Table 3. Based on the total binding energy a reranking was conducted which is depicted in Table 2. Showing the initial rank and the rank after readjusting based on the total interaction energy. Resulting in our top 3 candidates: 'ZINC001164979321', 'ZINC000620748806' and 'ZINC000514436632'.

Table 3: **Protein ligand interaction energy.** The average Coulomb and Lennard-Jones energies were computed for the simulations. The sum of both instances is the total interaction energy.

Interaction Energies			
Zinc ID	Coulomb Energy kJ/mol	Lennard-Jones kJ/mol	Total Interaction Energy kJ/mol
ZINC000019015192	-80.86 ± 16	-37.46 ± 8	-118.32 ± 17.89
ZINC000095523345	-2.78 ± 2.8	-5.40 ± 5.4	$-8.18 \pm 6,08$
ZINC000101042701	-0.45 ± 0.45	-0.57 ± 0.57	-1.02 ± 0.73
ZINC000104277568	-1.18 ± 1.1	-1.46 ± 1.4	-2.64 ± 1.78
ZINC000104309836	-88.10 ± 9.1	-86.87 ± 8.9	$-174,97 \pm 12.78$
ZINC000620748806	-185.15 ± 17	-127.63 ± 6.1	$-312,45 \pm 18,06$
ZINC001164872284	-48.39 ± 8.8	-102.95 ± 4.7	$-151,34 \pm 9.98$
ZINC001164979321	-301.87 ± 23	-112.65 ± 3.5	$-414,52 \pm 23,26$
ZINC000514436632	-198.37 ± 18	-110.15 ± 5.6	$-308,52 \pm 18.85$

4.5 Visualizing pocket stability by calculating the RMSD between protein and ligand

In order show that the ligands are stable and remain in the binding pocket, the RMSD between the protein and ligand was computed using internal *GROMACS* tools. The resulting '.xvg' files were visualized using the R package *'Peptides'* [Osório et al., 2015]. The RMSD plots of the reranked molecules are shown in Figure 3. Figure 3A; B; C show relatively high RMSD values compared to the rest of the model indicating big changes in the ligand poses. In contrast, the rest of the molecules mostly converged to a certain pose and hence, low RMSDs were observed, for instance our top 3 candidates after reranking Figure 3H; F;E show low RMSD values below 1 and are mostly constant. We can see high fluctuations in Figure 3F.

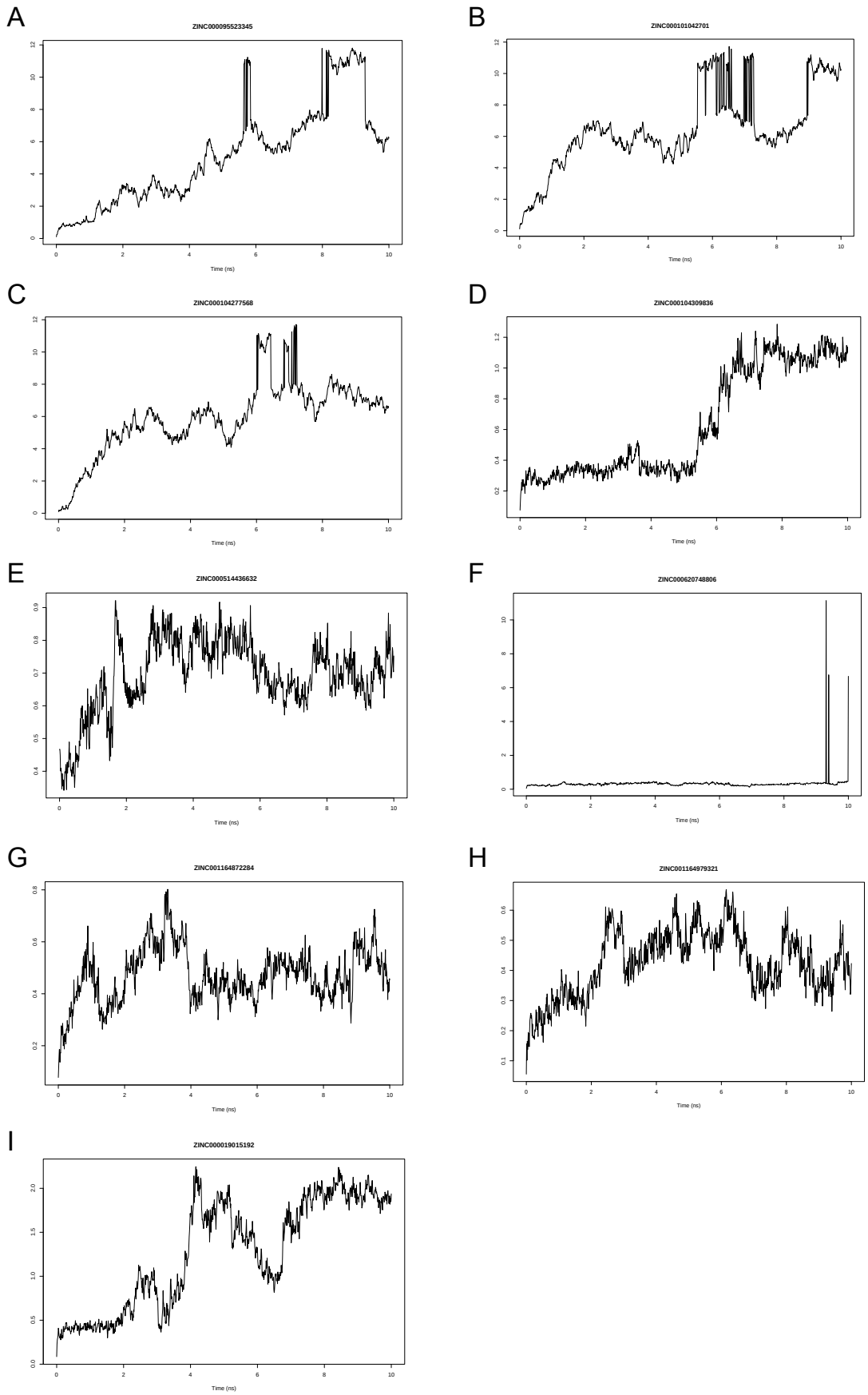


Figure 3: **Changes of the ligand relatively to NSP13.** A-I are annotated with the corresponding ZINC ID. The x axis is the time in ns and the y axis depicts the RMSD, meaning the relative change of the ligand in regard to nsp13.

4.6 AutoGrow 4 Results

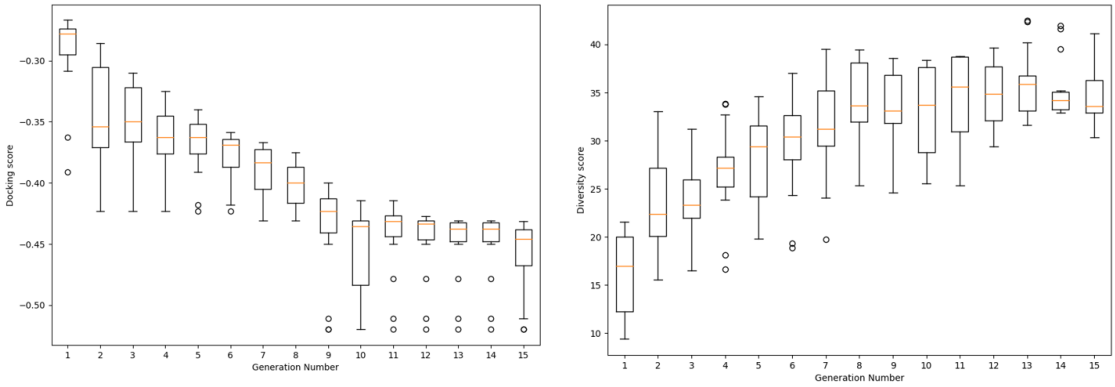


Figure 4: **Docking and diversity scores across AutoGrow 4 generations.**

Figure 4 shows how the docking scores and diversity scores change across generations. There is a visible improvement across the first 9 generations, and then the scores plateau. That indicates that 15 generation rounds could be excessive for the generation of the molecules and the results from earlier populations are worth exploring.

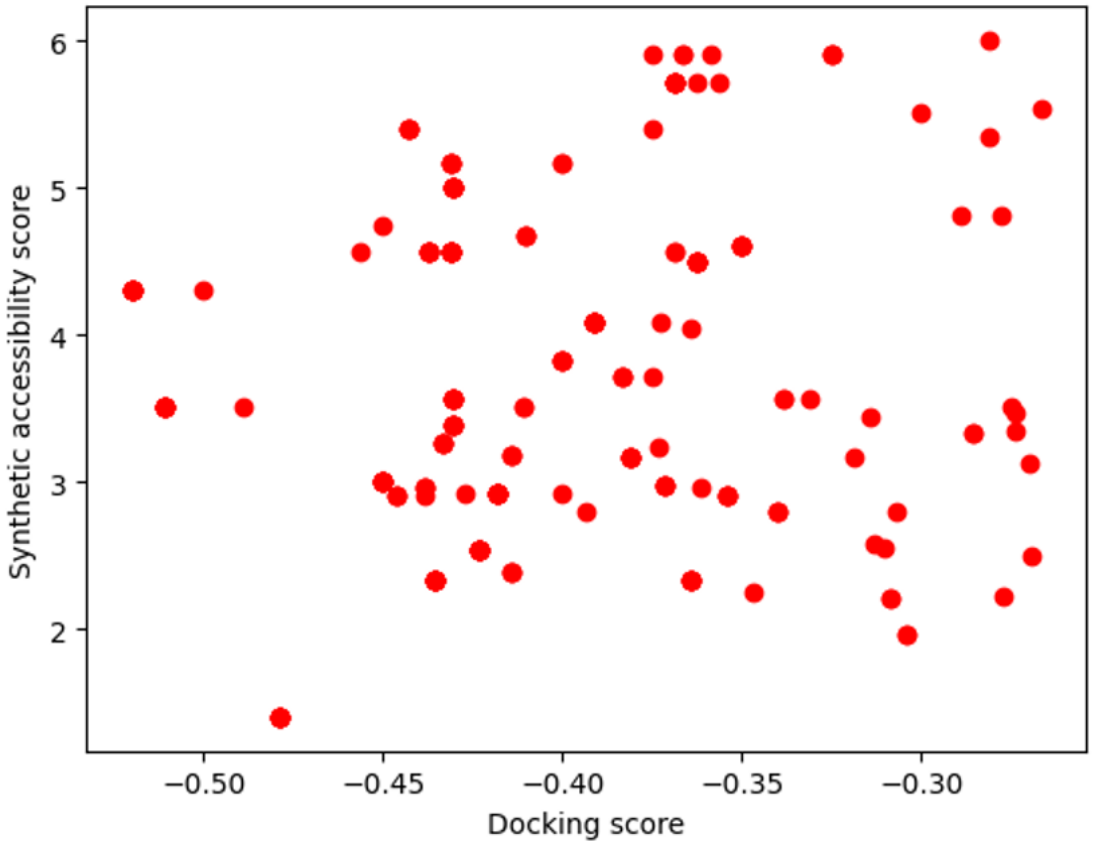


Figure 5: **Synthetic Accessibility Score on the top scoring molecules from the AutoGrow 4 against their respective Docking score.**

Figure 5 demonstrates the distribution of Synthetic Accessibility Scores (SAS) against the docking scores. There is no visible correlation, which is expected, as the best docking molecules are not neccessarily the most difficult to synthesize. It is also a good indication, that among best performers there are molecules that still can be rather easily synthesized. We explore the lower left quandrant with SAS less than 4 and the docking score less than -0.4 to find the possible new molecules.

Interestingly enough, among the generated molecules there are several that already have an identity, which are not among the provided in the input to AutoGrow 4. However, we are more interested in the newly generated molecules. The top scoring one with the least SAS score are visualized in Figure 6.

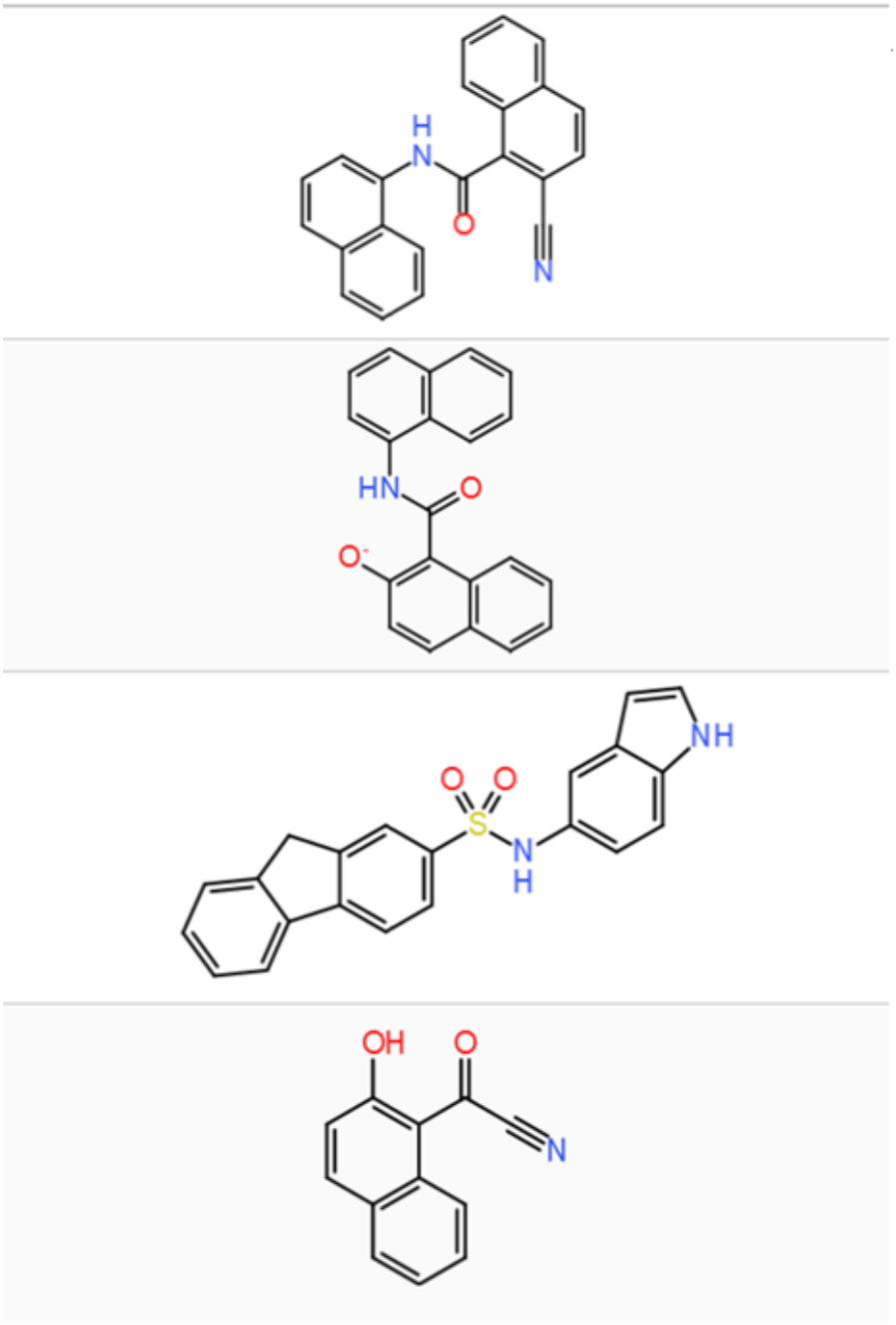


Figure 6: **Top generated molecules without known identity.**

5 Discussion and Outlook

The ligands with a binding affinity $< -10\frac{kcal}{mol}$ generally seem interesting for further consideration. Our Docking and ligand optimization results, like any other Docking results, should be computationally validated before being biologically validated in a further step. This should be done with different structures and the same docking tool and with different Docking tools. One tool we consider for further Docking runs in the future is the Glide docking tool [Halgren et al., 2004] and the

EnzyDock [Das et al., 2019] docking tool. Furthermore it would be interesting to screen a lot more molecules than we did in the scope of this project. Tools like *Vina-GPU* [Tang et al., 2022] and Gnina [McNutt et al., 2021] might help accelerating docking using GPUs.

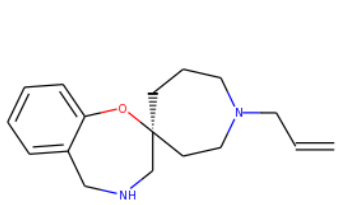
The MD simulation have been used as a tool to rerank or reanalyse a small fraction of molecules which include the best docking candidates (see Table rerank). *GROMACS* was used with default settings. Only 10 ns simulations were conducted but usually 100 or 150 ns are used [Hassab et al., 2022]. Both points can be optimized for our pipeline to ensure coherent results.

Many steps have been manually conducted in this workflow, for instance the '.pdb' file cleaning or the preparation of the docking results. Automating these workflows could make this approach more feasible. Another factor is the computational constraint, although *GROMACS* can be run on GPU, it still takes some time. In order to optimize this process, different tools can be benchmarked, comparing the comparability and speed. For instance, *Desmond* [6] could be tested.

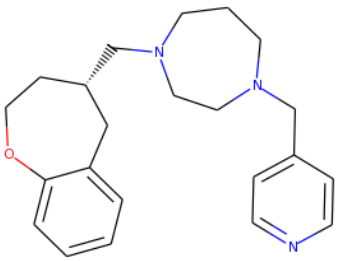
In general, our top 9 molecules have been successfully analyzed using docking and MD simulations, resulting in top 3 molecules that could be considered for further investigation.

[6] `https://www.schrodinger.com/products/desmond`

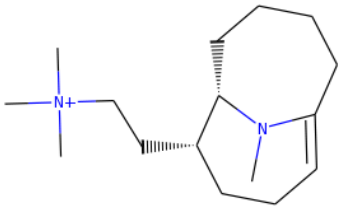
6 Supplementary Material



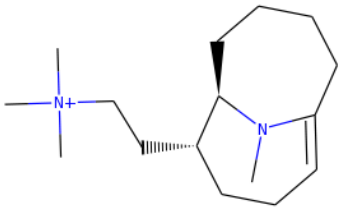
(a) ZINC000019015192



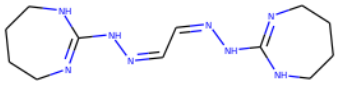
(b) ZINC000095523345



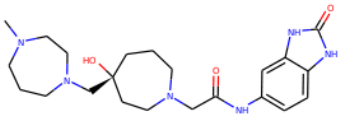
(c) ZINC000101042701



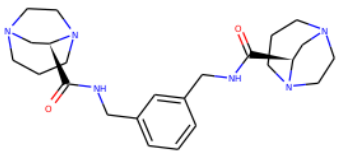
(d) ZINC000104277568



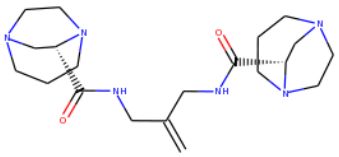
(e) ZINC000104309836



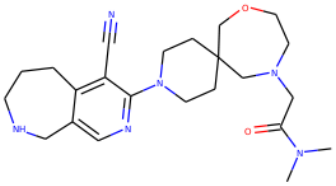
(f) ZINC000257281912



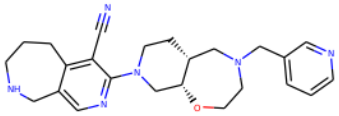
(g) ZINC000514436632



(h) ZINC000620748806

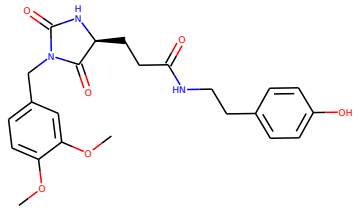


(i) ZINC001164872284

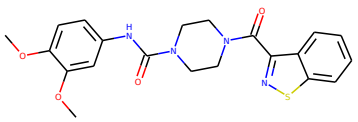


(j) ZINC001164979321

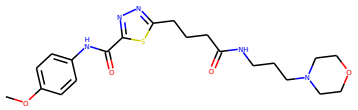
Figure 7: Molecule structures of the top ten molecules from our ZINC15-dataset subset after Docking with *AutoDock Vina* using the *Vinardo* scoring function.



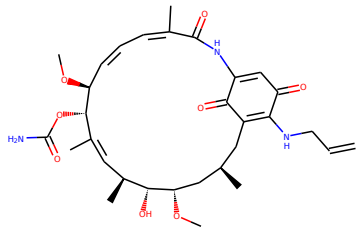
(a) EOS117



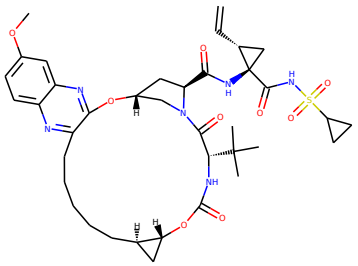
(b) EOS905



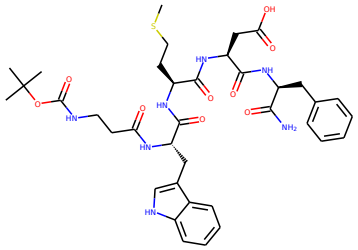
(c) EOS1709



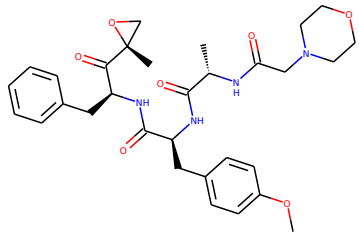
(d) EOS100609



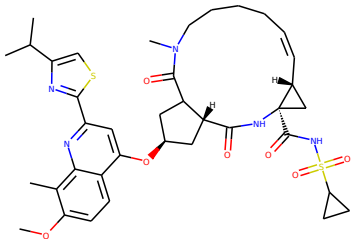
(e) EOS100851



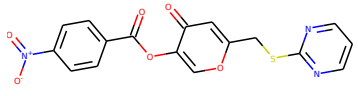
(f) EOS100853



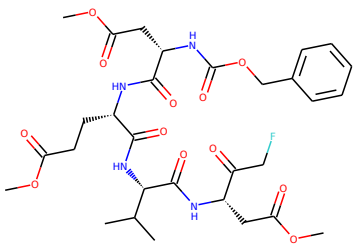
(g) EOS101170



(h) EOS101513

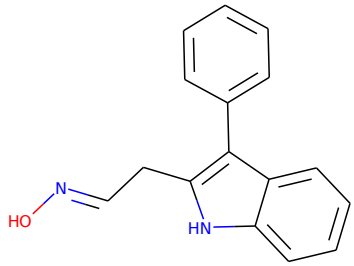


(i) EOS101554

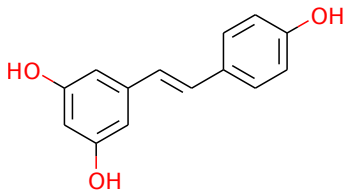


(j) EOS101596

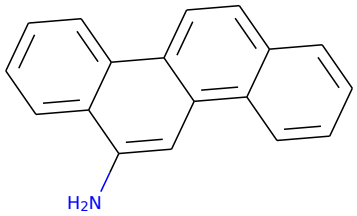
Figure 8: Molecule structures of the top ten molecules from The Sorbonne University Team from their EOS-database subset.



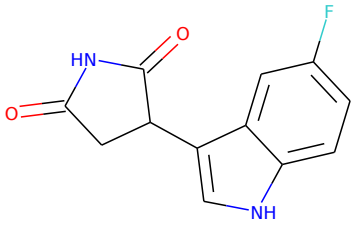
(a) EOS2253



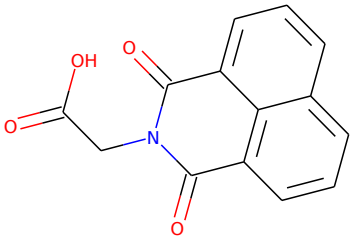
(b) EOS98620



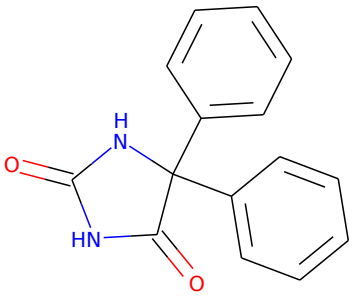
(c) EOS100134



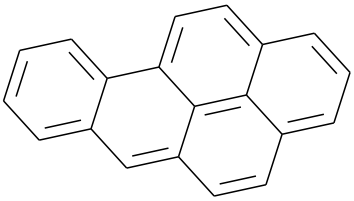
(d) EOS100433



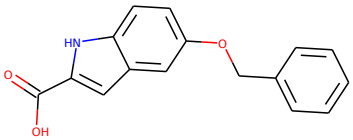
(e) EOS100710



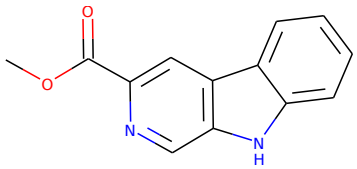
(f) EOS100899



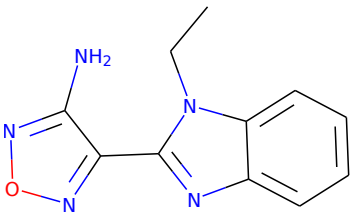
(g) EOS101630



(h) EOS102211



(i) EOS102407



(j) EOS102462

Figure 9: Molecule structures of the top ten molecules from the Milan University Team from their EOS-database subset.

References

[Al-Karmalawy et al., 2021] Al-Karmalawy, A. A., Dahab, M. A., Metwaly, A. M., Elhady, S. S., Elkaeed, E. B., Eissa, I. H., and Darwish, K. M. (2021). Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the hACE2 receptor. *Frontiers in Chemistry*, 9.

[Bekker et al., 1993] Bekker, H., Berendsen, H. J. C., Dijkstra, E., Achterop, S., Vondrumen, R., Vanderspoel, D., Sijbers, A., Keegstra, H., and Renardus, M. (1993). Gromacs - a parallel computer for molecular-dynamics simulations.

[Chen et al., 2022] Chen, J., Wang, Q., Malone, B., Llewellyn, E., Pechersky, Y., Maruthi, K., Eng, E. T., Perry, J. K., Campbell, E. A., Shaw, D. E., and Darst, S. A. (2022). Ensemble cryo-em reveals conformational states of the nsp13 helicase in the sars-cov-2 helicase replication–transcription complex. *Nature Structural & Molecular Biology*, 29:250–260.

[da Silva and Vranken, 2012] da Silva, A. W. S. and Vranken, W. F. (2012). Acpype - antechamber python parser interface. *BMC Research Notes*, 5:367 – 367.

[Das et al., 2019] Das, S., Shimshi, M., Raz, K., Eliaz, N. N., Mhashal, A. R., Ansbacher, T., and Major, D. T. (2019). EnzyDock: Protein–ligand docking of multiple reactive states along a reaction coordinate in enzymes. *Journal of Chemical Theory and Computation*, 15(9):5116–5134.

[Eastman et al., 2013] Eastman, P. K., Friedrichs, M. S., Chodera, J. D., Radmer, R. J., Bruns, C. M., Ku, J. P., Beauchamp, K. A., Lane, T. J., Wang, L., Shukla, D., Tye, T., Houston, M., Stich, T. J., Klein, C., Shirts, M. R., and Pande, V. S. (2013). Openmm 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of chemical theory and computation*, 9 1:461–469.

[Eberhardt et al., 2021] Eberhardt, J., Santos-Martins, D., Tillack, A. F., and 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):3891–3898.

[Forli et al., 2016] Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., and Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols*, 11(5):905–919.

[Guedes et al., 2018] Guedes, I. A., Pereira, F. S. S., and Dardenne, L. E. (2018). Empirical scoring functions for structure-based virtual screening: Applications, critical aspects, and challenges. *Frontiers in Pharmacology*, 9.

[Guilloux et al., 2009] Guilloux, V. L., Schmidtke, P., and Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, 10(1).

[Halgren et al., 2004] Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. enrichment factors in database screening. *Journal of Medicinal Chemistry*, 47(7):1750–1759.

[Handoko et al., 2012] Handoko, S. D., Ouyang, X., Su, C. T. T., Kwoh, C. K., and Ong, Y. S. (2012). Quickvina: accelerating autodock vina using gradient-based heuristics for global optimization. *IEEE/ACM transactions on computational biology and bioinformatics*, 9(5):1266–1272.

[Hanwell et al., 2012] Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4:17 – 17.

[Hassab et al., 2022] Hassab, M. A. E., Eldehna, W. M., Al-Rashood, S. T., Alharbi, A., Eskandrani, R. O., Alkahtani, H. M., Elkaeed, E. B., and Abou-Seri, S. M. (2022). Multi-stage structure-based virtual screening approach towards identification of potential SARS-CoV-2 NSP13 helicase inhibitors . *Journal of Enzyme Inhibition and Medicinal Chemistry*, 37(1):563–572.

[Hornak et al., 2006] Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. E., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure*, 65.

[Humphrey et al., 1996] Humphrey, W. F., Dalke, A., and Schulten, K. (1996). Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14 1:33–8, 27–8.

[Jendele et al., 2019] Jendele, L., Krivak, R., Skoda, P., Novotny, M., and Hoksza, D. (2019). PrankWeb: a web server for ligand binding site prediction and visualization. *Nucleic Acids Research*, 47(W1):W345–W349.

[Jiménez-Luna et al., 2021] Jiménez-Luna, J., Grisoni, F., Weskamp, N., and Schneider, G. (2021). Artificial intelligence in drug discovery: Recent advances and future perspectives. *Expert opinion on drug discovery*, 16(9):949–959.

[Joshi et al., 2020] Joshi, T., Joshi, T., Pundir, H., Sharma, P., Mathpal, S., and Chandra, S. (2020). Predictive modeling by deep learning, virtual screening and molecular dynamics study of natural compounds against SARS-CoV-2 main protease. *Journal of Biomolecular Structure and Dynamics*, 39(17):6728–6746.

[Lemkul, 2019] Lemkul, J. A. (2019). From proteins to perturbed hamiltonians: A suite of tutorials for the gromacs-2018 molecular simulation package [article v1.0]. *Living Journal of Computational Molecular Science*.

[Luttens et al., 2022] Luttens, A., Gullberg, H., Abdurakhmanov, E., Vo, D. D., Akaberi, D., Talibov, V. O., Nekhotiaeva, N., Vangeel, L., Jonghe, S. D., Jochmans, D., Krambrich, J., Tas, A., Lundgren, B., Gravenfors, Y., Craig, A. J., Atilaw, Y., Sandström, A., Moodie, L. W. K., Lundkvist, Å., van Hemert, M. J., Neyts, J., Lennerstrand, J., Kihlberg, J., Sandberg, K., Danielson, U. H., and Carlsson, J. (2022). Ultralarge virtual screening identifies SARS-CoV-2 main protease inhibitors with broad-spectrum activity against coronaviruses. *Journal of the American Chemical Society*, 144(7):2905–2920.

[McNutt et al., 2021] McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., Sunseri, J., and Koes, D. R. (2021). GNINA 1.0: molecular docking with deep learning. *Journal of Cheminformatics*, 13(1).

[Meyers et al., 2021] Meyers, J., Fabian, B., and Brown, N. (2021). De novo molecular design and generative models. *Drug Discovery Today*, 26(11):2707–2715.

[Newman et al., 2021] Newman, J. A., Douangamath, A., Yadzani, S., Yosaatmadja, Y., Aimon, A., Brandão-Neto, J., Dunnett, L., Gorrie-stone, T., Skyner, R., Fearon, D., Schapira, M., von Delft, F., and Gileadi, O. (2021). Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nature Communications*, 12(1).

[O'Boyle et al., 2011] O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011). Open babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1):1–14.

[Osório et al., 2015] Osório, D., Rondón-Villarreal, P., and Torres, R. (2015). Peptides: A package for data mining of antimicrobial peptides. *R J.*, 7:4.

[Pedretti et al., 2002] Pedretti, A., Villa, L., and Vistoli, G. (2002). Vega: a versatile program to convert, handle and visualize molecular structure on windows-based pcs. *Journal of molecular graphics & modelling*, 21 1:47–9.

[Pu et al., 2019] Pu, L., Naderi, M., Liu, T., Wu, H.-C., Mukhopadhyay, S., and Brylinski, M. (2019). eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates. *BMC Pharmacology and Toxicology*, 20(1).

[Pushpakom et al., 2018] Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., and Pirmohamed, M. (2018). Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1):41–58.

[Quiroga and Villarreal, 2016] Quiroga, R. and Villarreal, M. A. (2016). Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLOS ONE*, 11(5):e0155183.

[Raubenolt et al., 2022] Raubenolt, B. A., Islam, N., Summa, C. M., and Rick, S. W. (2022). Molecular dynamics simulations of the flexibility and inhibition of sars-cov-2 nsp 13 helicase. *Journal of Molecular Graphics & Modelling*, 112:108122 – 108122.

[S, 2010] S, D. (2010). Mgltools.

[Spiegel and Durrant, 2020] Spiegel, J. O. and Durrant, J. D. (2020). Autogrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *Journal of cheminformatics*, 12(1):1–16.

[Sterling and Irwin, 2015] Sterling, T. and Irwin, J. J. (2015). ZINC 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337.

[Tang et al., 2022] Tang, S., Chen, R., Lin, M., Lin, Q., Zhu, Y., Ding, J., Hu, H., Ling, M., and Wu, J. (2022). Accelerating AutoDock vina with GPUs. *Molecules*, 27(9):3041.

[White et al., 2020] White, M. A., Lin, W., and Cheng, X. (2020). Discovery of COVID-19 inhibitors targeting the SARS-CoV-2 nsp13 helicase. *The Journal of Physical Chemistry Letters*, 11(21):9144–9151.