SORBONNE UNIVERSITÉ

European Bioinformatics Masters Network

MEET-U Report

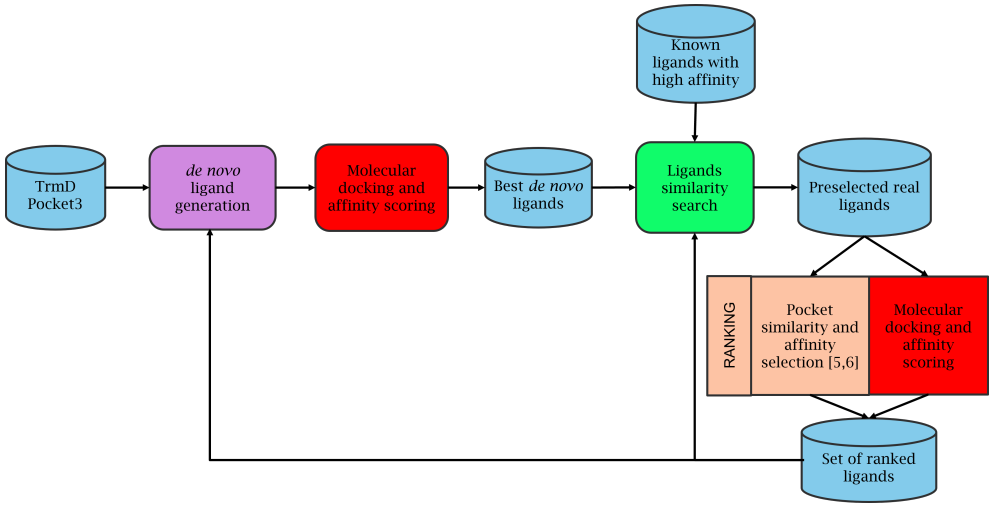
Bettiati M., Boitel L., Hacquard T., Jatiere S., Pautet F.

March 1, 2025

Abstract

The bacterial tRNA methyl transferase has for several years been a target in drug discovery due to its critical role in bacterial growth and absence in eukaryotes' cells. Several experimental and computational compound screening have been realized, to no concrete avail. In the following report, we raise a concern regarding the selection of the compound being screened and the underused public databases in recent publications. Following this, we outline here a pipeline with the goal of efficiently screen through such databases to quickly evaluate the druggability of small ligand molecules. We developed novel approaches leveraging cutting cutting-edge tools to select promising ligands, evaluate their docking ability on TrmD and their potential in inhibiting it.

Graphical Abstract



Code and Data availability:
https://github.com/meetU-MasterStudents/2024-2025$_{T}eam1-SU/$

1 Introduction

Drug-resistant bacteria are becoming more and more frequent. As a result from the extensive use of antibiotics causing the selection of resistant strains, many diseases are consecutively becoming challenging to treat [23, 3]. This stresses the importance of finding alternatives. In that regard, researchers face a double challenge. On the one hand, they must find a suitable candidate protein and/or pathway to inhibit the bacteria's growth and/or survival. On the other hand, the medicine must be harmless to humans, while not degraded by our metabolism or immune system. Hence, wide screening methods have been developed, first experimental (for example with Minimal Inhibitory Concentration (MIC) tests coupled with cytotoxicity tests), and more recently computational, with the rise of computational protein-docking in drug-discovery.

One of the most promising bacterial is the tRNA Methyl-Transferase TrmD. It encompasses several of the aspects mentioned previously, being vital for bacterial growth and not shared with eukaryotes [9]. Consecutively, finding an inhibitor of the TrmD could lead to broad pharmaceutical applications, given that it is found in most pathogenic bacteria. The TrmD has been extensively studied, and research have been led to identify viable inhibitors [8, 4, 26, 7].

The TrmD is a homo-dimer (hereafter denominated chain A and chain B) containing two actives sites able to receive its main cofactor, the S-adenosy-methyl (AdoMet). Activity cannot take place unless both monomers are joined together, although while both may bind the AdoMet substrate, only chain B possess the ability for the tRNA substrate binding [10], and only one AdoMet can be bound at a time. The TrmD catalyzes the methylation of the N^1 of the Guanine 37 in tRNAs. The relevance of such activity is explained by the use of programmed nucleotide frame-shifting in mRNA translation by ribosomes in bacteria, as a mean of gene expression regulation. As detailed in [9], this will be deleterious for protein's activity, and the 37G tRNAs' methylation allow to avoid such mistakes. Consecutively, the association of the TrmD activity with tRNAs prone to frame-shifts has been found to divide by ten the number of frame-shifts [6], thus justifying why it is crucial for bacterial growth.

Several sites could be considered for docking-based inhibition of the TrmD. The tRNA binding sites, AdoMet pockets, or the Mg^{2+}. However, in this work and for this quick overview we will focus on the AdoMet pocket site. In 2019 Zhong, Wenhe *et al.* experimentally screened more than 100,000 compounds, and about 30 new compounds with TrmD inhibition function were found [26]. In 2020, about a thousand compounds were screened after resolving the TrmD structure by X-ray crystallography [19], and in 2023, 200 compounds were designed, and screened *in silico* prior an *in vivo* evaluation. It is however our understanding that in public databases such as PubChem or Plinder resides dozens of millions of compounds, which were never indicated of being tested or screened. Hence, we propose to develop a pipeline that would focus precisely on that. The following report is the outline of a pipeline whose main focus is set on the rapid ligands screening from public databases. We developed and tried new methods for ligands selection coupled with docking steps to use to their full extent the massive compound libraries at our disposal.

2 Methods

2.1 Similarity search

The number of chemical component with an experimental affinity to TrmD [25](100 000 molecules) is 3 orders of magnitude lower than the number of components in pubchem [1] (71 millions molecules). Starting from a molecule with a known affinity[25], we compute its neighborhood using a fingerprint similarity [16]. This simple notion of similarity can be intuitively understood as selecting molecules which share a certain number of substructures (i.e. similar chemical functions, number of bonds,...). To refine this initial set, we use the notion of flexible common substructure: for two molecules, we compute the common substructure with the maximum number of atoms (we are allowing 1 atom and 1 bond of error). At the end, for each reference ligand with an experimental affinity we are associating a neighborhood set of about 200 ligands.

2.1.1 Fingerprint search

The first step to compare our reference ligand to the database is to compute the fingerprint of every molecules. The fingerprint is a function from the space of molecules M to the space of binary words 2^n. A given molecule is described by its signature, which encodes the presence or absence of given substructures into a binary word s(m). The similarity between 2 words x and y is then measured with the Tanimoto coefficient:

$$\tau(x, y) = \frac{\#(x \cap y)}{\#(x \cup y)} \tag{1}$$

In this equation, # computes the number of 1 in the word, $\cap$ is the bitwise and and $\cup$ is the bitwise or. The similarity between 2 molecules m and n is the similarity of their signatures:

$$\tau(m, n) = \tau(s(m), (s(n)) \tag{2}$$

By selecting the closest 200 molecules for this notion of similarity, we get a set of ligands similar the ones known to bind with the TrmD. However, the signature based similarity is not enough to discriminate between molecules that appear similar and those that actually behave similarly. This is why we refine the initial search using a flexible Maximum Common substructure (fMCS) search[21].

2.1.2 Similarity tree

The fMCS search models a molecule as a planar multigraph. Given 2 molecules, it computes the flexible maximum common substructure that is a common substructure of both molecules with an error of at most 1 atom an 1 liaison with maximum number of atoms. We then compute the associated Tanimoto coefficient as a similarity. Using our starting reference ligand and the set given by fingerprint search we construct a set using the algorithm 1. In practice we are constructing a tree by recording which n in the set S has been used to add a molecule de S_1.

Data: ref_lignand is the initial reference ligand
Data: Q is the set of ligands returned by the fingerprint querry
Result: $S \subset Q$ is the subset of ligand which are of interest
$S \leftarrow \{\text{ref\_ligand}\}$;
for $i = 1$ **to** 2 **do**
  $S_1 \leftarrow \{m \in Q \mid \exists n \in S : \tau(n, m) > 0.4\}$;
  $S \leftarrow S \cup S_1$;
end

Algorithm 1: Similarity tree

2.2 De novo Ligand Generation

2.2.1 Approach

Our similarity search strategy relies on the small set of known ligands for their affinity to TrmD. To enrich our search space, we integrated to our pipeline another way to obtain ligands of interest: *de novo* generation via a diffusion model.

By using a machine learning model trained on structures of various protein-ligand complexes, we can massively generate new ligands that were not observed previously with a potential good affinity to our target pocket, the SAM active site of TrmD.
The main limitation of this approach is usually that those ligands are unlikely to be chemically synthetically. Our hypothesis is that our *de novo* ligands with the best docking affinity to our site will be good reference points for a similarity search among known existing ligands - which would enable us to derive interesting new candidates among our database, without the downside of uncertain chemical viability.

In addition to its complementarity with our similarity search, our generative model also benefits from the rest of our pipeline: we use rigid-body docking, pocket similarity scoring, and flexible docking to filter among the thousands of potential *de novo* ligands we generate. Only the best ones are then used for our similarity search.

2.2.2 Model architecture

Several machine learning architectures have been explored previously for targeted ligand generation, using diverse architectures, such as variational auto-encoders [15], graph neural networks [18], and diffusion models [5].
We decided to focus on the diffusion model presented in [5]. Our decision was guided by three factors: it outperforms other methods on benchmarks such as on the dataset presented in [13], it includes chemical synthetic accessibility as one of the target objectives of the training (which makes it more likely to yield interesting similarity search starting points), and diffusion models in general are a very active research area with potential for important future improvements.

The PILOT architecture uses a diffusion process: the atom positions, types charges, and bond types from a protein-ligand training example are noised, and the model's objective is to reconstruct the ground truth. In practice, a diffusion forward trajectory is built by iteratively applying Gaussian noise to the ligand, and the learning is done on the different steps of this trajectory. The model itself consists of 12 layers of message-passing.

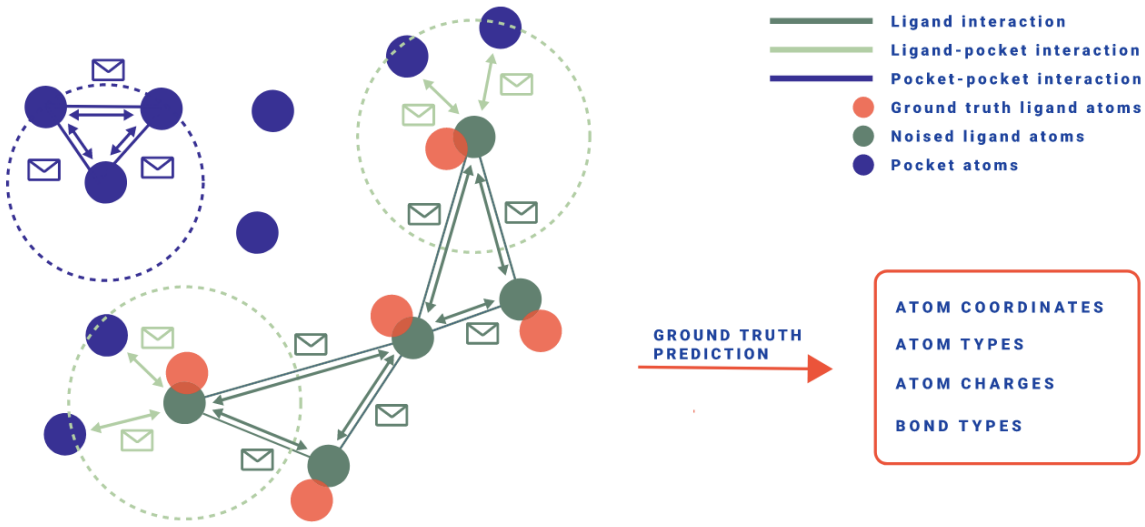


Figure 1: Figure from [5] - network of attention-weighed message passing among atoms, used by the model to try to reconstruct the ligand ground truth from the noisy output of forward diffusion.

2.2.3 PILOT model training

On top of its ability to reconstruct the physical parameters of a noisy ligand, the model also regresses on other loss functions that are combined together (multi-objective training). The authors of [5] introduce a measure of synthetic accessibility and a docking score with the target pocket. Those are combined together to guide the model, not only towards ligands that fit well the pocket, but towards ligands that are more realistic and likely to bind to it.

In practice, we ran the training on the Institut Francais de Bioinformatique (IFB) cluster. Due to the cluster's contraints, we only had one A100 GPU to train on. We had to modify some implementation details to allow it to run on our hardware. Among other things, we modified the PyTorch Lightning

training loop logic to skip PDB files that didn't fit in our GRAM, initially causing our training runs to crash early on. We trained only on the CrossDocked dataset, skipping pre-training as we didn't have the resources to handle the massive amounts of data required. We also had to patch some pipelining to allow the code to run with the libraries we had access to, and with the CrossDocked dataset version we used.

We intended to use the TrmD-specific docking data generated by the rest of our pipeline to do transfer learning, and further train our model (after freezing most layers weights), to specialize it to TrmD ligand generation. However, we did not have time to implement this due to the time spent on training the base model. This approach would have allowed us to get more specialized ligands, but might have led to some overfitting as we only would have a limited amount to high quality TrmD ligands to present to the model. We still think it may be worth exploring, since our pipeline would allow us to generate new structures of known ligands docked in silico to TrmD.

2.3 Pocket Generation and Similarity

The docking methods usually involve a measure of the affinity of the ligand bound to the protein. This measure however is often either a minimization of free energy (i.e., not a formal expression of the thermodynamic propensity of the docking reaction), or a scoring based on machine learning, and thus lacking biological or physical meaning, like in Gnina, whom we will use hereafter. Hence, we propose another way of estimating the binding affinity of a ligand for a protein, in addition to this traditional measure. The key idea resides in the generation of an "ideal" pocket around the indicated ligand. The generated pocket then docks the ligand, and is compared to the pocket shape obtained from the traditional methods of docking. Consecutively, we derive a notion of distance between the generated pocket and the natural pocket for a tested ligand. This distance is compared relatively to the reference distance between the generated pocket of Adomet with TrmD respect to the pocket where Adomet has been docked. This boils down to determining if the pocket of the screened ligands able to inhibit the TrmD are closest to their ideal pocket than the original ligand, here the AdoMet for TrmD.

In this endeavor, the main tools we used are PocketGen, recently published in 2024 [24]. PocketGen is a deep-learning generative model, whose baseline is the ligand docked in the natural pocket of the protein. It starts by isolating the pocket of the protein that interacts with the ligand. Then, the model identifies the interactions of residues with the ligand, and then dives deeper in those interactions, to the atom level. This part leverages the concept of attention, i.e. the incorporation of context information from both the amino-acid and ligand atoms to design interactions on both molecular and atomistic scales. This allows to identify features that can be updated and improved, which is done afterward, with a sequence refinment step. This step relies on a protein Language Model, thus using the Transformer architecture, similarly to Foldseek [20]. It updates the amino-acids in the sequence following the layout designed by the previous part, resulting in a novel pocket sequence with improved binding affinity.

In order to establish pockets similarity levels, we used the tool EPOCS, described in a preprint on bioArxiv [17]. It uses the popular deep-learning architecture ESM-2 to construct a metric of Cross-similarity between the pockets required to be compared, which we then leverage for the purpose described above, and compute a pairwise distance matrix based on the Euclidean metric.

2.4 Ligand scoring and docking

Once we have a set of promising ligands, we evaluate them using Gnina [14]. Gnina is a molecular docking program with integrated support for scoring and optimizing ligands using convolutional neural networks. Extending the AutoDock framework, it uses convolutional neural networks (CNNs) trained on protein-ligand complexes to evaluate docking poses and predict binding affinities. This machine learning approach allows Gnina to capture complex molecular interactions that traditional scoring functions might miss.

Like most docking tools, Gnina start by sampling the conformational space. This conformational space is huge, as both the receptor and the ligand may be flexible, allowing for each molecule to adjust its shape to the influence of the other. It is too computationally expensive to enable the whole target to

be flexible. However, Gnina allows for some protein flexibility by defining flexible sidechains, while leaving the backbone fixed. To reduce the conformational space, it is also recommended to target the docking onto a known pocket, rather than to perform docking on the whole protein. In our case, we focus on the binding site of the AdoMet in the high-resolution crystal structures of TrmD in a binary complex with AdoMet, characterized in the pdb 1uak obtained by Hyung Jun Ahn et al. [2]. The docking is thus performed in the box defined by the AdoMet ligand.

Gnina samples the conformational space via a Monte Carlo chain. Following the completion of Monte Carlo sampling, the sampled conformations are evaluated with a scoring function and only the best conformations are retained for further refinement. Refinement shifts the ligand pose to a local energy minimum using the gradients of the scoring function. After the ligand pose has been refined, the final affinities and scores are calculated for refined the pose.

The scoring function is an empirical measure of the free energy which predicts the strength of interaction between the ligand and the protein, wherelower binding energy indicates higher predicted affinity. Gnina allows to use several different empirical scoring functions such as Vina or Vinardo score, but also implements multiple pretrained CNN scores. However, the refinement step is much faster when using a a functional formulation, rather than a grid approximation as performed by the CNN. Therefore, the CNN is only used to evaluate and compare the final positions.

For each ligand, we thus obtain multiple binding poses, as well as a score measuring the affinity of the ligand to the TrmD in each position. Those are sorted by affinity and converted into the output format sdf using OpenBabel.

3 Results

3.1 Similarity tree

We tested 20 ligand of the paper. 4 were not in the database and failed. The 16 remaining ligands gave us 2665 potentials ligands. The main issue that we had with the similarity method is that the references ligands tends to be unique in the database. The closest match does not seem to be that similar 2. As a consequence, we did not find molecules very close to the ligand with only subtle differences. Typically the molecules found were widely different from the reference. While this result is interesting in itself, it will also make the results of docking less certain, as well as casting doubts on the validity of this approach. This is illustrated by Figure 3, where it three clusters are found to exist around the reference molecule. This means that while the molecules in those cluster should be similar to each other, they are only so close to the molecule supposed to find a neighborhood of ligands whose druggability would be tested..

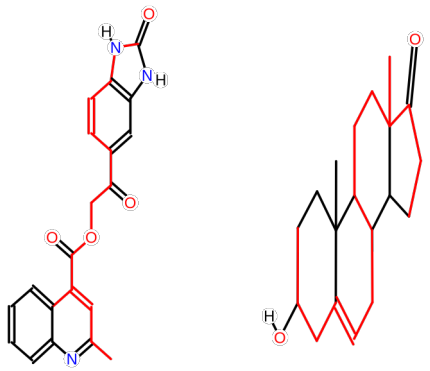


Figure 2: Comparaison between a known lignad and the closest hit using our similarity researcgh algorithm.

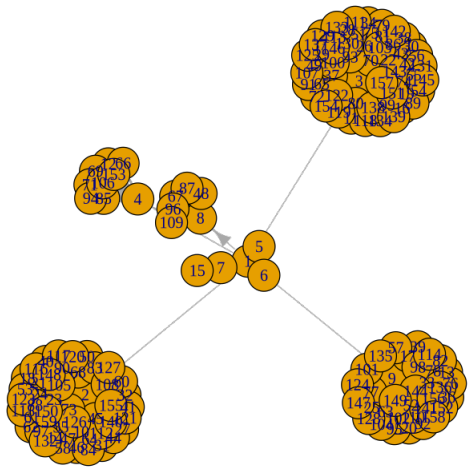


Figure 3: Tree resulting of our research of known ligands. Each node is a potential ligand

3.2 *De novo* Ligand Generation

3.2.1 Generation

As discussed in methods, we had to adapt our training process to our available ressources. We did not have the time to finish training the network - the authors of [5] recommend using 300 epochs of training, but we only could run 22. We still saw some improvements over our training time in terms of validation loss, and we hope to have a much better model before the final presentation in Milan, so we can update our results.

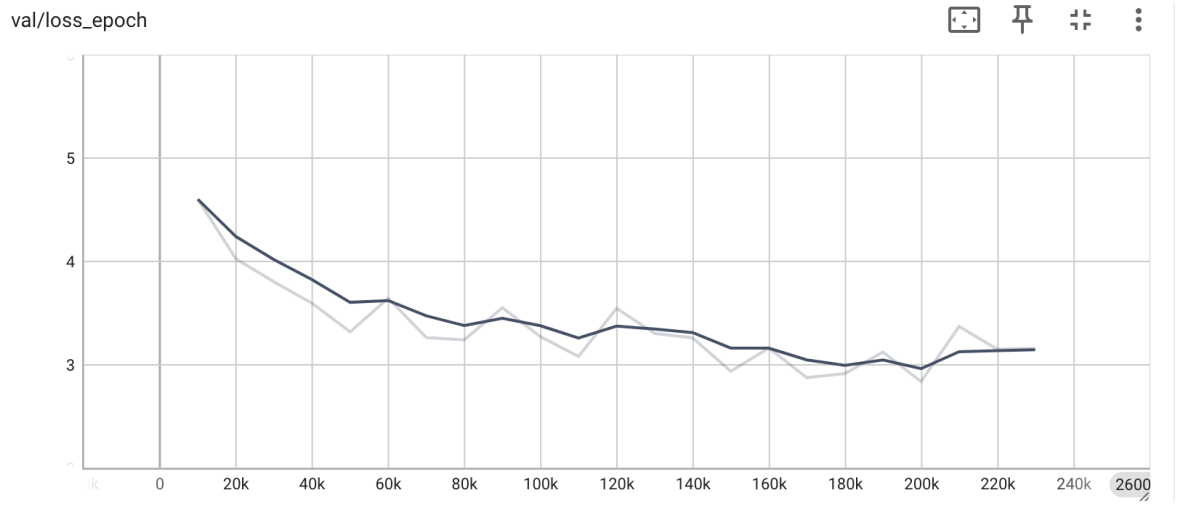


Figure 4: Validation loss evolution over 22 epochs of training on CrossDocked data.

We generated 28 ligands as a proof of concept, and visualized them in PyMol.

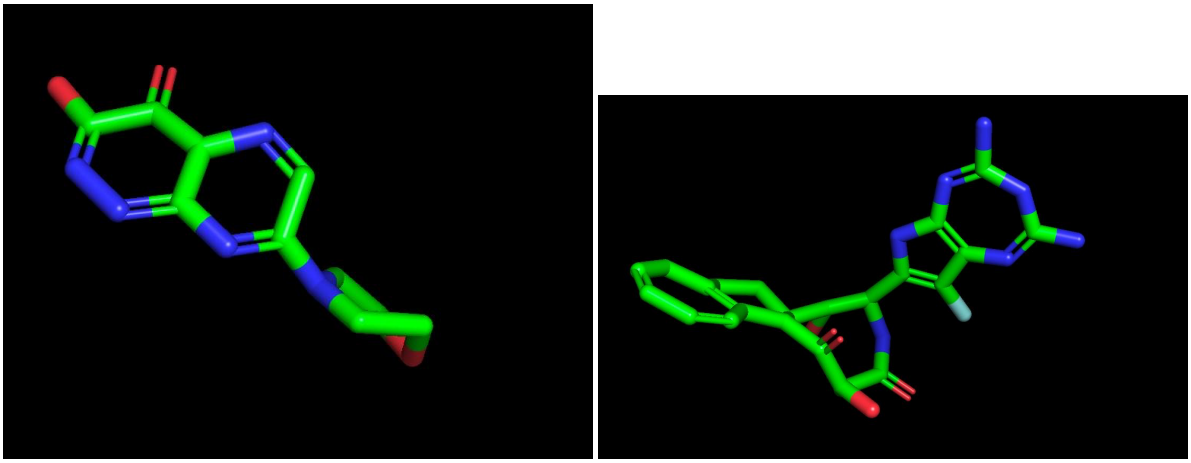


Figure 5: Two de novo ligands generated with an early model.

3.2.2 Use as a basis for similarity search

We ran a similarity search using one of our de novo ligands as target. We were able to find some matches, which confirms the viability of our pipeline. Once again, we are waiting for the model's training to progress in the next few days, so we can scale up on the most interesting de novo ligands we will have generated.

3.2.3 Evaluation of the de novo ligands

Gnina was used to dock the thus created ligands and evaluate their affinity to the TrmD. Their affinity score was compared to the predicted affinity of the AdoMet for the TrmD. From the 28 de novo generated ligands, 9 got a higher affinity. Figure 3.2.3 shows the insertion of the best scoring ligand into the pocket of the TrmD.

This shows that the pipeline works. The artificial ligands can be docked onto the TrmD and may even outperform its canonical ligand, the AdoMet.

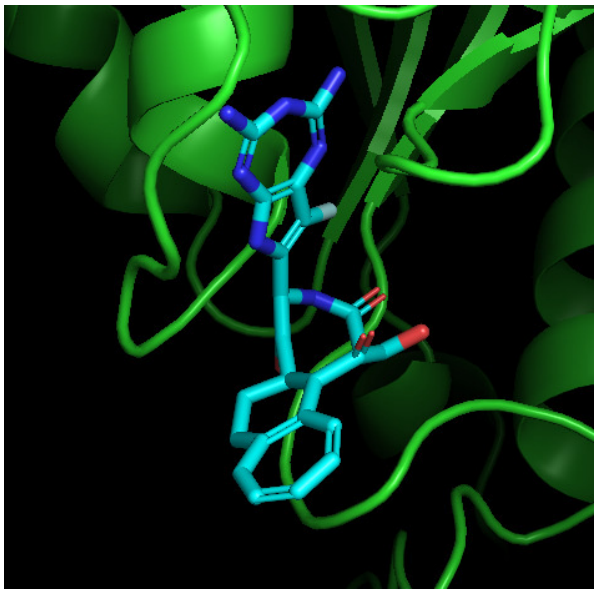


Figure 6: De novo generated ligand docked into TrmD (higher affinity than SAM, even if it is made with an early model and likely not synthetizable)

3.3 Pocket Generation and Similarity

As a proof of concept for this method of relative affinity measurement, only six structure were tested. One structure is the TrmD complexed with AdoMet, its natural ligand and can be found under UniProt code 5WYQ. Another is the folded TrmD associated with the sinefugin, under UniProt code 4YPX. Sinefungin is a molecule known for its antibiotic properties, and a structural analog to the AdoMet. The other four are other ligands we found through our pipeline and tested. The main structural difference between the two molecules lies in the absence of sulfur-activated methyl in the Sinefungin, respect to Adomet, thus preventing the methylation reaction from taking place. Figure 7 shows the Euclidean distance matrix across pockets, and displays the distance to their generated "ideal" pocket. Consecutively, it stands out that the natural AdoMet ligand pocket (5wyq_lig1) is farther to its generated pocket than the Sinefungin. Following our reasoning detailed in our Methods, this would mean that their is a greater affinity of the TrmD for Sinefungin than AdoMet. In addition, all other four ligands are farther farther from both their ideal pocket, and TrmD's natural one. However, verifying this hypothesis would require dedicated experimental and statistical testing, which we do not have time to do. As a consequence we will rely on literature results to corroborate our reasoning. Sinefungin has been extensively used as a TrmD competitive inhibitor [22, 12, 11]. More specifically, in 2011 and 2019, Jaroensuk *et al.* and Lahoud *et al.* both showed that Sinefungin concentration as low as 10 to 4 time smaller than AdoMet's were enough for competitive inhibition, suggesting a higher affinity for Sinefungin than Adomets on the TrmD part [11, 12]. In addition, all other four ligands are known to have lower affinity for TrmD than the Adomet, further corroborating our initial hypothesis. Nonetheless, an eventual statistical relationship has yet to be derived as this is only a toy example. Hence, while this result does corroborate the validity of our approach, it only merely hints at the potential of this approach. As far as we know, such procedure has never been described in the published literature, and we shall develop our insights and leads for improvements and testing in the Discussion section.

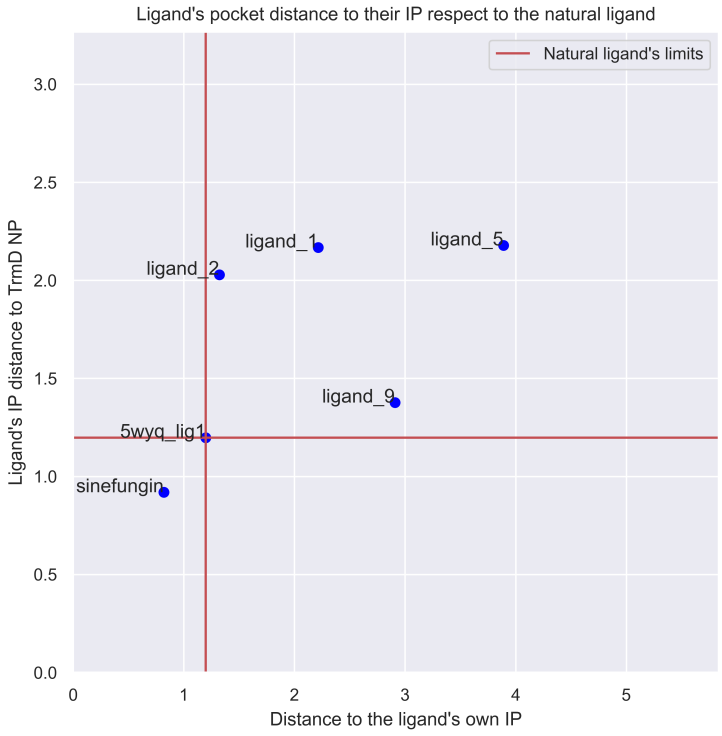


Figure 7: Scatter plot of ligands closeness to both their own Ideal Pocket (IP), and their ideal pocket respect to the TrmD's Natural Pocket (NP).

4 Discussion

4.1 Similarity tree

Using the notion of similarity we are able to expend on the existing experimental dataset. However, we can only expand on molecules presents in the PubChem dataset. We have been limited by the public PubChem API which limits the number of queries to fixed number, thus 50, 100, 200 and 1000 at a time. Consecutively, the processivity of the approach is limited. Future work could aim to access other databases to mitigate this issue. Furthermore, the parallelization of flexible maximal common substructure was unsuccessful due to R inability to handle multi-threading. Launching multiple parallel processes was surprisingly slower than the single threaded implementation. Thus, resolving those issues would enhance yet again the processivity of this similarity search, enabling a much more efficient use of the under-used available databases. Finally, explaining the cause of the isolation of some ligands found to be of interest in previous studies could allow a more fruitful neighborhood search of ligands with similar structure as those given as input.

4.2 De novo Ligand Generation

Even if we were limited in training time, our early model allows us to discover interesting compounds, some with better docking performance than SAM, and to then look for similar real ligands using our similarity search pipeline. This already allowed us to find interesting ligands.

More training time should allow us to scale up our results and have bigger batches of good candidate ligands before the final presentation. We won't have time to explore model specialization to TrmD but we expect it to be a promising avenue for future research.

4.3 Docking

By analyzing binding energies and docking poses, we expect the best candidate ligands to have a consistently higher binding affinity than the others. The integration of Gnina allows an efficient screening and scoring of the ligands obtained from the similarity searches or from the de novo generation. It is a highly flexible tool, providing a variety of functional scores, pretained CNN along with the possibility for user defined scores. It also allows to specify the function for each step of the process (sampling, refinement and final evaluation).

The computational efficiency and flexibility of Gnina make it well-suited for handling large datasets. Although Gnina provides valuable insights, it has limitations. Its scoring function may not accurately reflect experimental binding affinities due to simplified energy models. For now, we used the default empirical scoring for the docking and the CNN only for the final evaluation as it requires more time. However we could take advantage of parallelization to use more complex and hopefully more accurate scoring functions. Furthermore, we did not look into protein flexibility yet. For now, ligands were docked into the rigid pocket suited for AdoMet, making the flexibility an interesting path to explore.

4.4 Pocket Generation and Similarity

The novel approach we present here of comparing "ideal" docking pockets with "natural" docked pockets suffers from a number of bias that cannot be ignored. It is unclear for example whether Pocketgen's pockets would be actually viable in living beings, which would challenge the notion of "ideal pocket" we defined previously. In addition, due to the time constraints of the assignment, a proper statistical analysis of the process efficiency could not be realized, and only one example was produced. To complete this picture, and for the same reasons, the models were not retrained, and only the parameters prepared by the authors were used, resulting in possible biases. Further work would then have to start by engineering proper workflow and analysis on a statistically significant sample of ligands supposed to inhibit a protein activity, with the natural substrate as the reference. In addition, another focus would be the formal definition of the notion of "ideal pocket", and try to establish a correlation between the ligand affinity and distance to the ideal pocket. Another angle that could be

interesting to develop are links between kinetic enzymatic activity and distance between generated and ideal pockets.

5 Conclusion and Perspectives

Our objective in this project was to develop an approach that would allow for a better use of the available databases, and as a consequence, propose screening processes. Enabling an efficient ligand pre-selection enhances the success rate of the docking, thus enabling faster and more successful the finding of small molecules as drug candidates.

Every step of the approach we designed and hypothesized were put in place, and whilst the quantity of results does lack, each individual step is working, and produces the expected results on simple examples. We designed novel approaches that hint at success, especially the ligand similarity search. The pocket-generation affinity estimation has yet to prove its efficiency in more realistic scenarios. In addition, we combined several other methods of ligand proposal, such as the *de novo* generation of ligand molecules, although proper testing has yet to be achieved, just like for the whole pipeline. In that regard, we successfully implemented in our pipeline cutting edge docking tools such as AutoDock Vina and Gnina, leveraging their processing efficiency in the optic, again, to take advantage of the enormous size of public database of small ligand. Finally, we made our pipeline a circular system, with the idea of using promising ligands as tuning parameters for both the generation model and the similarity search.

Our most direct perspectives are to finalize the remaining missing links in our pipeline, and give it a trial run at producing drug candidates inhibiting TrmD, followed by the fine-tuning of every steps and scaling of the entire pipeline.

References

[1] URL: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72.

[2] Hyung Jun Ahn et al. "Crystal structure of tRNA (m1G37) methyltransferase: insights into tRNA recognition". In: *The EMBO journal* (2003).

[3] Aleksandra Baran, Aleksandra Kwiatkowska, and Leszek Potocki. "Antibiotics and bacterial resistance—a short story of an endless arms race". In: *International Journal of Molecular Sciences* 24.6 (2023), p. 5777.

[4] Thomas Christian et al. "Distinct origins of tRNA (m1G37) methyltransferase". In: *Journal of molecular biology* 339.4 (2004), pp. 707–719.

[5] Julian Cremer et al. *PILOT: Equivariant diffusion for pocket conditioned de novo ligand generation with multi-objective guidance via importance sampling*. 2024. DOI: 10.48550/ARXIV.2405.14925. URL: https://arxiv.org/abs/2405.14925.

[6] Howard B Gamper et al. "Maintenance of protein synthesis reading frame by EF-P and m1G37-tRNA". In: *Nature communications* 6.1 (2015), p. 7226.

[7] Pamela J Hill et al. "Selective inhibitors of bacterial t-RNA-(N1G37) methyltransferase (TrmD) that demonstrate novel ordering of the lid domain". In: *Journal of medicinal chemistry* 56.18 (2013), pp. 7278–7288.

[8] Hiroyuki Hori. "Transfer RNA methyltransferases with a SpoU-TrmD (SPOUT) fold and their modified nucleosides in tRNA". In: *Biomolecules* 7.1 (2017), p. 23.

[9] Ya-Ming Hou et al. "TrmD: a methyl transferase for tRNA methylation with m1G37". In: *The Enzymes* 41 (2017), pp. 89–115.

[10] Takuhiro Ito et al. "Structural basis for methyl-donor–dependent and sequence-specific binding to tRNA substrates by knotted methyltransferase TrmD". In: *Proceedings of the National Academy of Sciences* 112.31 (2015), E4197–E4205.

[11] Juthamas Jaroensuk et al. "Crystal structure and catalytic mechanism of the essential m1G37 tRNA methyltransferase TrmD from Pseudomonas aeruginosa". In: *Rna* 25.11 (2019), pp. 1481–1496.

[12] Georges Lahoud et al. "Differentiating analogous tRNA methyltransferases by fragments of the methyl donor". In: *RNA* 17.7 (2011), pp. 1236–1246.

[13] Shitong Luo et al. "A 3D Generative Model for Structure-Based Drug Design". In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 6229–6239. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/314450613369e0ee72d0da7f6fee773c-Paper.pdf.

[14] Andrew T McNutt et al. "GNINA 1.0: molecular docking with deep learning". In: *Journal of cheminformatics* 13.1 (2021), p. 43.

[15] Nhat Khang Ngo and Truong Son Hy. "Target-aware Variational Auto-encoders for Ligand Generation with Multimodal Protein Representation Learning". In: (Aug. 2023). DOI: 10.1101/2023.08.10.552868. URL: http://dx.doi.org/10.1101/2023.08.10.552868.

[16] Nina Nikolova and Joanna Jaworska. "Approaches to measure chemical similarity–a review". In: *QSAR & Combinatorial Science* 22.9-10 (2003), pp. 1006–1026.

[17] Tugce Oruc et al. "Mapping the space of protein binding sites with sequence-based protein language models". In: *bioRxiv* (2024), pp. 2024–07.

[18] Xingang Peng et al. *Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets*. 2022. DOI: 10.48550/ARXIV.2205.07249. URL: https://arxiv.org/abs/2205.07249.

[19] Sherine E Thomas et al. "Fragment-based discovery of a new class of inhibitors targeting mycobacterial tRNA modification". In: *Nucleic Acids Research* 48.14 (2020), pp. 8099–8112.

[20] Michel Van Kempen et al. "Fast and accurate protein structure search with Foldseek". In: *Nature biotechnology* 42.2 (2024), pp. 243–246.

[21] Yan Wang et al. "fmcsR: mismatch tolerant maximum common substructure searching in R". In: *Bioinformatics* 29.21 (Aug. 2013), pp. 2792–2794. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt475. eprint: https://academic.oup.com/bioinformatics/article-pdf/29/21/2792/50746515/bioinformatics_29_21_2792.pdf. URL: https://doi.org/10.1093/bioinformatics/btt475.

[22] Joseph M Watts, J Gabruzsk, and Walter M Holmes. "Ligand-mediated anticodon conformational changes occur during tRNA methylation by a TrmD methyltransferase". In: *Biochemistry* 44.17 (2005), pp. 6629–6639.

[23] Chunming Xu et al. "A review of current bacterial resistance to antibiotics in food animals". In: *Frontiers in Microbiology* 13 (2022), p. 822689.

[24] Zaixi Zhang et al. "Efficient generation of protein pockets with PocketGen". In: *Nature Machine Intelligence* (2024), pp. 1–14.

[25] Wenhe Zhong et al. "Targeting the Bacterial Epitranscriptome for Antibiotic Development: Discovery of Novel tRNA-(N1G37) Methyltransferase (TrmD) Inhibitors". In: *ACS Infectious Diseases* 5.3 (2019). PMID: 30682246, pp. 326–335. DOI: 10.1021/acsinfecdis.8b00275. eprint: https://doi.org/10.1021/acsinfecdis.8b00275. URL: https://doi.org/10.1021/acsinfecdis.8b00275.

[26] Wenhe Zhong et al. "Targeting the bacterial epitranscriptome for antibiotic development: discovery of novel tRNA-(N1G37) methyltransferase (TrmD) inhibitors". In: *ACS infectious diseases* 5.3 (2019), pp. 326–335.