

REPORT: SOLVING TrmD BINDING PROBLEM USING EVOLUTION INSPIRED ALGORITHM

PREPRINT, COMPILED JANUARY 25, 2025

Anna Gregušová  Tomáš Jelínek  Tomáš Preisler  Aneta Tranová  Marek Tobiáš 

Faculty of Science, Charles University

Team Prague 2

Supervisor: Mgr. Marian Novotný, Ph.D. 

ABSTRACT

The growing problem of antibiotic resistance makes it crucial to find new antibacterial targets and treatments. Bacterial tRNA (guanine-N1)-methyltransferase TrmD is an attractive drug target due to its role in tRNA modification and its structural divergence from its human counterpart, Trm5. This study presents a computational approach for TrmD inhibitor discovery, integrating virtual screening, molecular docking, and an evolutionary optimization algorithm, EvoFLOPA. Initially, FDA-approved molecules were screened against three TrmD binding sites (AdoMet pocket, tRNA binding site, and dimerization interface) to establish a benchmark. EvoFLOPA, leveraging SELFIES molecular representation and UniDock docking, optimized candidate compounds based on binding affinity, synthetic accessibility (SA), and quantitative estimate of drug-likeness (QED). This pipeline identified Compound YF and its optimized derivative YFOH, both exhibiting high binding affinity and favorable pharmacokinetic and pharmacodynamic properties. Further rescoring with RTMScore and SwissDock validated their potential as TrmD inhibitors. This study highlights a promising computational strategy for developing novel TrmD inhibitors with high affinity and favorable drug properties.

1 Introduction

The relentless rise of antibiotic-resistant bacteria poses a critical global health challenge, demanding urgent efforts to discover and develop novel antibacterial therapies [1]. Among the promising targets for new antibiotics is tRNA (guanine-N1)-methyltransferase TrmD, an important bacterial enzyme responsible for catalyzing the methylation of guanine at position 37 (m^1G37) in tRNA [2]. This modification is crucial for preventing +1 frameshifts during translation. TrmD's bacterial specificity, resulting from its structural and functional divergence from its human counterpart, Trm5, makes it an attractive target for developing species-selective antimicrobials, minimizing potential off-target effects in human cells [3].

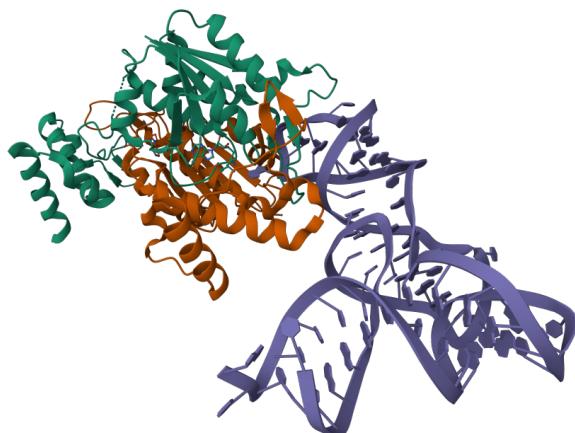


Figure 1: PDB ID: 4YVI complex. TrmD protein's homodimer chains (green and orange), tRNA (shown in blue)

Traditional high-throughput screening approaches are often time-consuming and resource-intensive. *In silico* methods, particularly virtual screening and molecular docking, offer a powerful and efficient alternative for identifying potential drug candidates[4]. Evolutionary algorithms, inspired by natural selection, provide a further sophisticated approach for lead optimization and *de novo* molecular design, capable of navigating vast chemical spaces and identifying compounds with desired properties[5].

In this study, we introduce a computational pipeline for TrmD inhibitor discovery, integrating virtual screening of FDA-approved molecules, molecular docking, and EvoFLOPA, a novel evolutionary algorithm designed for molecular optimization. EvoFLOPA employs SELFIES [6] molecular representation, UniDock for docking, and a multiobjective scoring function that incorporates binding affinity, synthetic accessibility, and drug-likeness. We apply this approach to identify promising TrmD inhibitors, notably Compound YF and its derivative YFOH, which exhibit high binding affinity and improved pharmacokinetic and pharmacodynamic properties. Further rescore and analysis confirm their potential as lead compounds, demonstrating the effectiveness of computational evolution in antibacterial drug discovery.

2 Methods

Our computational workflow consisted of two primary phases: **(a)** docking of FDA-approved molecules for benchmarking our virtual screening pipeline, and **(b)** application of the EvoFLOPA evolutionary algorithm for *de novo* lead optimization. For molecular docking, we selected the *H. influenzae* TrmD resolved protein structure (PDB ID: **4YVG**) [7] due to its high resolution and relevance as a model organism for antibiotic development. *H. influenzae* is a common target in antibacterial drug discovery, making this structure particularly suitable for studying TrmD inhibitors with potential broad spectrum activity.

Using **P2Rank** [8], we identified multiple potential binding sites on TrmD and selected the two most promising ones based on their accessibility and druggability: the *AdoMet binding pocket* and the *tRNA binding site*.

With further literature analysis, we have identified a spot on the protein that when mutated inhibited TrmD's activity [9], possibly due to blockage of its dimerization, that is, because TrmD protein's activity is attributed to its homodimer form, so we have labeled another docking site, which we call the *dimerization interface*.

2.1 FDA Approved Molecules Docking and Benchmarking

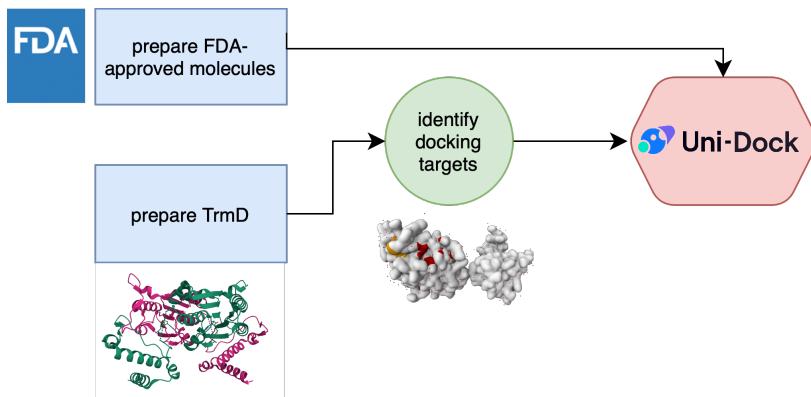


Figure 2: Workflow of our virtual screening method using FDA-approved molecules.

To establish a baseline and validate our docking protocol, we performed virtual screening using a library of FDA-approved molecules. The workflow is depicted in Figure 2.

- **Preparation of FDA-approved molecules:** The FDA-approved molecule library was prepared using RDKit [10], an open-source cheminformatics toolkit. This involved:
 - **Sanitization:** Ensuring chemical validity and correcting any inconsistencies in molecular representations.
 - **Protonation:** Adding hydrogen atoms and assigning appropriate protonation states at physiological pH.
 - **Charge Assignment:** Assigning Gasteiger charges for accurate electrostatic calculations during docking.
 - **Component Selection:** Taking the largest molecular component of each FDA-approved drug to overcome docking problems with multiple-component structures (removing free ions, etc.)
- **Preparation of TrmD protein:** The crystal structure of TrmD from *Haemophilus influenzae* (PDB ID: 4YVG) [7] was used as a receptor for docking studies. Protein preparation was carried out using standard protocols:

- **Water Removal:** Removing water molecules (HOH residues) to prevent interference with ligand docking in the binding sites.
- **Hydrogen Addition:** Adding hydrogen atoms to ensure correct protein protonation.
- **Identification of Docking Targets:** Three potential binding sites on TrmD were identified using P2Rank [8], a machine learning-based tool for ligand binding site prediction:
 - **AdoMet Pocket:** The binding site for the cofactor S-adenosylmethionine (SAM), crucial for TrmD’s enzymatic activity.
 - **tRNA Binding Site:** The region where TrmD interacts with its tRNA substrate.
 - **Dimerization Site:** The interface between the two TrmD monomers, potentially offering allosteric inhibition opportunities.

Table 1: Binding Site Specifics

Docking Site	Coordinates (x, y, z) Å	Size (x, y, z) Å
AdoMet Pocket	(47.8, 9.4, 14.1)	(20, 20, 20)
tRNA Binding Site	(29.5, 9.4, 26.2)	(20, 20, 20)
Dimerization Site	(47.0, -12.7, -9.5)	(15, 15, 15)

- **Molecular Docking:** Docking simulations were performed using UniDock [11], a GPU-accelerated docking program known for its speed and accuracy in large-scale virtual screening. We utilized a Vinardo scoring function, offered by UniDock, to estimate binding affinity. We used *balanced* UniDock screening option (with exhaustiveness=384, max_step=40, seed=42).
- **Benchmarking:** The docking scores obtained for FDA-approved molecules were analyzed to establish a baseline for binding affinity and to assess the performance of our docking protocol in identifying potential TrmD binders.

2.2 EvoFLOPA Evolutionary Algorithm

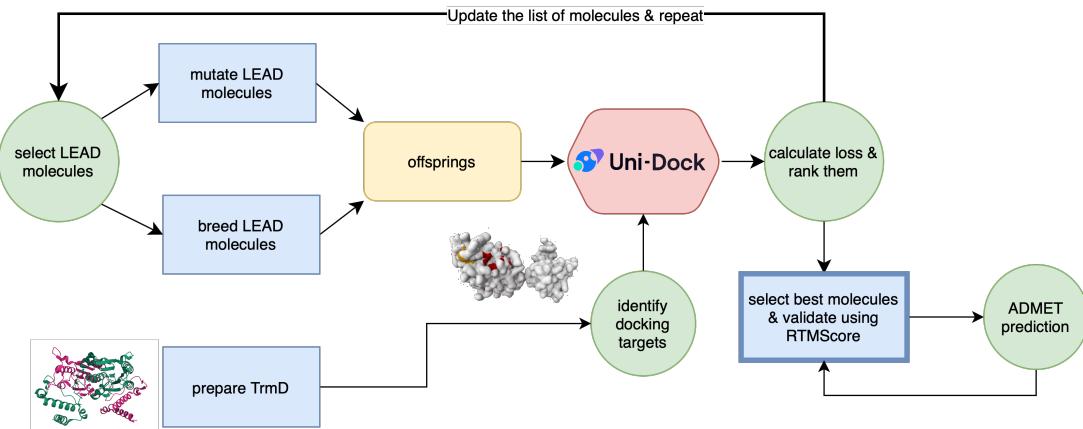


Figure 3: Flowchart of EvoFLOPA evolutionary algorithm workflow.

To actively optimize lead molecules inside the AdoMet binding pocket and explore novel chemical space, we developed EvoFLOPA, an evolutionary algorithm. The workflow of EvoFLOPA is illustrated in Figure 3. For further information, refer to its GitHub repository (<https://github.com/Desperadus/EvoFLOPA/>)

- **Starting Molecules (LEAD Molecules):** We selected a set of known TrmD inhibitors, specifically Compounds 21, 23, 35, 36, and 37 reported by Wilkinson et al. [12], which demonstrated promising binding affinity in previous studies, as our initial "LEAD" molecules for the evolutionary process.
- **Molecular Representation (SELFIES):** EvoFLOPA employs SELFIES [6] (Simplified Molecular-Input Line-Entry System) strings to represent molecules. SELFIES offers a robust and chemically valid representation, allowing for direct manipulation of molecular graphs through string operations while ensuring syntactically correct and chemically plausible molecules are generated throughout the evolutionary process.

- **Mutation:** A modified STONED algorithm [13] was implemented for introducing mutations into the SELFIES strings of parent molecules. Three types of mutations were applied with equal probability (33% each):
 - **Atom Addition:** Adding a new atom to the molecule.
 - **Atom Deletion:** Removing an existing atom from the molecule.
 - **Atom Replacement:** Replacing an atom with a different atom type. Mutations were performed at the token level of the SELFIES string to maintain chemical validity.
- **Breeding:** EvoFLOPA incorporates a breeding step where genetic material from two parent molecules is combined to create offspring. This process involved:
 - **Levenshtein Distance Calculation:** Computing the Levenshtein distance (edit distance) between the SELFIES strings of two parent molecules.
 - **Edit Path Generation:** Determining the shortest edit path to transform one SELFIES string into another.
 - **Offspring Creation:** Generating offspring molecules by applying a series of edits along the calculated edit path, effectively combining features from both parent molecules.
- **Docking and Scoring:** Generated molecules were docked into the AdoMet binding site of TrmD using UniDock, as described in Section 2.1. The binding affinity was assessed using the Vinardo scoring function.
- **Loss Function:** A multi-objective loss function was designed to guide the evolutionary process towards molecules with desirable properties:

$$L = W_{SA} * (1 - SA/10) + W_{QED} * QED - W_{Dock} * DockingScore \quad (1)$$

Where:

- **SA (Synthetic Accessibility):** A score estimating the ease of chemical synthesis, calculated using RDKit. A lower SA score indicates easier synthesis.
- **QED (Quantitative Estimate of Drug-likeness):** A score predicting the drug-likeness of a molecule, also calculated using RDKit. A higher QED score suggests better drug-like properties.
- **DockingScore:** The Vinardo docking score obtained from UniDock. More negative scores indicate stronger predicted binding affinity.
- **W_{SA} , W_{QED} , W_{Dock} :** User-defined weights controlling the relative contribution of each term to the overall loss function. The loss function aims to be maximized, effectively balancing binding affinity, synthetic feasibility, and drug-likeness.
- **Iterative Optimization:** The evolutionary optimization process was iterative, proceeding through generations:
 - **Ranking and Selection:** Molecules in each generation were ranked based on their loss function scores.
 - **Top Molecule Selection:** The top n molecules (e.g., top 128) from each generation were selected and appended to a history list of the best molecules.
 - **Probabilistic Selection for Next Generation:** From the history list, a subset of molecules was probabilistically selected to seed the next generation. The probability of selection was determined by converting loss scores into probabilities using a softmax function with a temperature parameter (T). A lower temperature favors exploitation (selecting molecules with higher loss scores), while a higher temperature promotes exploration (allowing for selection of lower-scoring molecules).
 - **Mutation and Breeding:** The selected molecules underwent mutation and/or breeding to generate a new population of molecules for the next generation.
 - **Iteration:** This cycle of docking, scoring, selection, mutation, and breeding was repeated for a defined number of iterations (e.g., 500 or 1000).
- **Validation with RTMScore:** To further validate the binding poses and refine affinity predictions, the top-ranked molecules identified by EvoFLOPA were rescored using RTMScore [14], a deep learning-based scoring function. RTMScore learns from experimental protein-ligand complexes and calculates likelihoods of observed residue-atom distances, offering improved accuracy and generalization compared to empirical scoring functions like Vinardo.
- **ADMET Prediction:** To assess potential drug-like properties beyond QED, ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) predictions were performed. Specifically, DILI (Drug-Induced Liver Injury) [15] prediction was conducted to evaluate potential toxicities.

2.2.1 RTMScore: A Residue-Atom Distance Likelihood Scoring Function

Recent advancements in machine learning (ML) have fueled the development of protein-ligand scoring functions (SFs). To address the limitations of existing approaches, Shen et al. introduced RTMScore, a novel SF based on a residue-atom distance likelihood potential and a graph transformer architecture.

RTMScore's key innovation is its use of a residue-based graph representation for proteins, where nodes represent residues, and edges define their interactions. Ligands are encoded with conventional atom and bond features. Multiple graph transformer layers learn representations for both protein and ligand, with information extracted by the transformers being concatenated and fed into a mixture density network (MDN). The MDN calculates the probability distribution of minimum distances between each residue and ligand atom, and the negative log-likelihood values from the MDN are combined to yield a statistical potential (i.e., score). This approach learns complex protein-ligand interactions without relying on predefined additive descriptors.

Performance and Comparison of RTMScore

RTMScore was rigorously evaluated on the CASF-2016 benchmark and has been shown to achieve a higher success rate compared to traditional approaches in docking tasks, with the top-ranked poses being within 2 Å RMSD of the native structure in approximately 93% of the complexes. Even with the inclusion of the crystal structure, this value becomes close to 97% (Figure 4). Furthermore, RTMScore shows superior performance in screening tasks, with high success rates in both forward and reverse screening tests, achieving top-1% enrichment factors of close to 30. These results position RTMScore favorably among other methods.

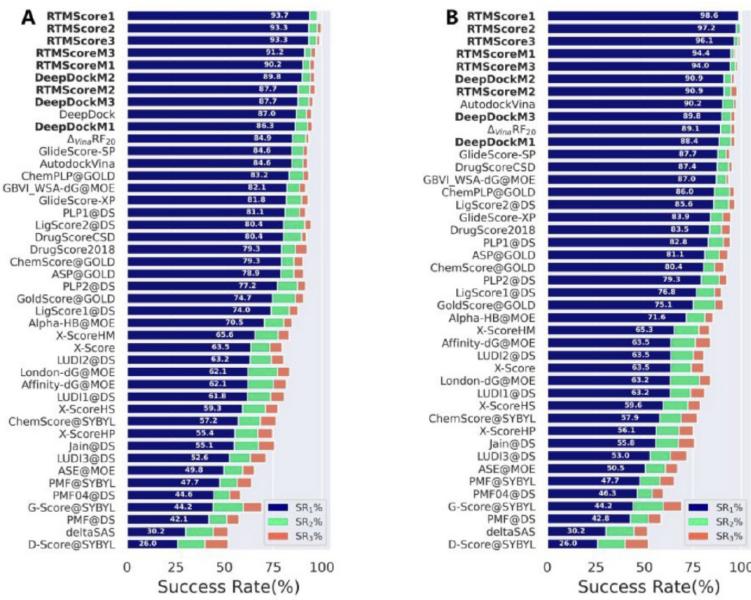


Figure 4: Performances of scoring functions on the CASF-2016 benchmark, including the docking powers in terms of (A) success rates excluding crystal poses and (B) success rate including crystal poses[14].

To confirm the robustness and generalization capabilities, the PDBbind-CrossDocked-Core data set and the DEKOIS2.0 and DUD-E datasets were also used for additional validation. Results show that RTMScore's excellent predictive abilities are maintained in the cross-docking scenario.

DeepDock (which also uses a distance likelihood potential) employs a mesh graph of the protein with relative Cartesian coordinates while RTMScore uses a residue graph. Consequently, DeepDock's performance is sensitive to protein rotations, while RTMScore's performance is improved by using graph transformer layers, making it more robust.

In addition to the benchmarks reported in the original RTMScore paper, we have also observed its superior performance in our own docking experiments. For example, when docking two different compounds, namely Cp23 and Cp21, into the AdoMet binding pocket, the traditional docking method, AutoDock Vina scores Cp21 as better (-9.834) than Cp23 (-8.698). This contradicts the experimental evidence. In contrast, RTMScore correctly distinguishes the binding preference, scoring Cp23 as superior (34.5) to Cp21 (33.5). This highlights the advantage of using RTMScore in scenarios where traditional docking methods might fail.

3 Results

3.1 Results of FDA-Approved Molecule Docking

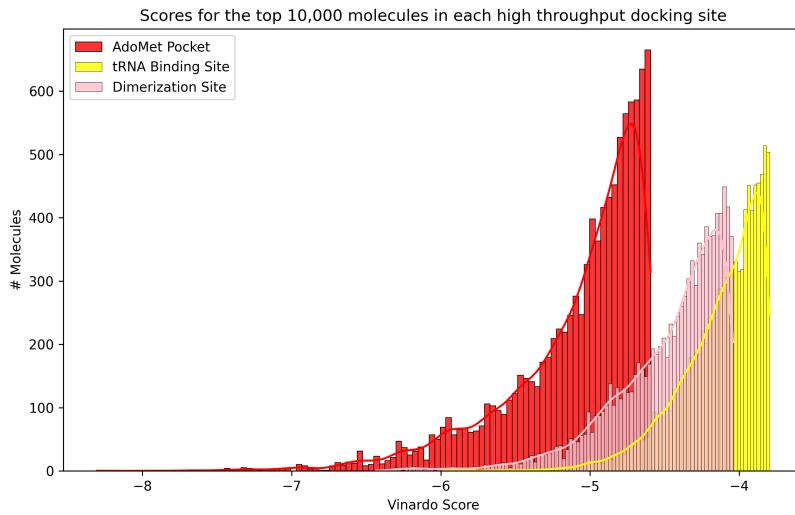


Figure 5: Histogram of docking score distributions for FDA-approved molecules across three TrmD binding sites. Only top-scoring 10,000 molecules (affinity-wise) from each binding site are displayed. Red distribution refers to AdoMet binding pocket, yellow - tRNA binding site, pink - dimerization site.

Docking of FDA-approved molecules against the AdoMet pocket, tRNA binding site, and dimerization site of TrmD provided a benchmark for our virtual screening approach. Figure 5 displays the distribution of Vinardo docking scores for the top 10,000 molecules screened against each site. More negative Vinardo scores indicate stronger predicted binding affinity.

The analysis revealed distinct score distributions across the three sites. The AdoMet pocket exhibited a range of scores with higher affinity compared to other sites, suggesting a greater potential to bind diverse molecules. The tRNA binding site and dimerization site showed distributions that shifted more towards positive values, potentially indicating more constrained binding preferences.

The top-ranked FDA-approved molecules for each site, based on Vinardo scores, were:

- **AdoMet Pocket:** *Glycerol phenylbutyrate* (Vinardo Score: -13.5 kcal/mol, RTMScore: 32.16)
- **tRNA Binding Site:** *Temoporfin* (Vinardo Score: -5.983 kcal/mol)
- **Dimerization Site:** *Cyclobenzaprine hydrochloride* (Vinardo Score: -6.433 kcal/mol)

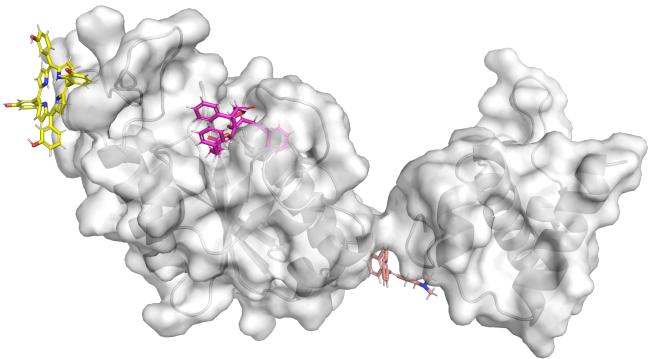


Figure 6: PyMOL visualization of 4YVG with top-rated docked FDA-approved molecules bound to each of the three identified TrmD binding sites

Figure 6 illustrates the predicted binding poses of these top molecules within their respective TrmD binding sites. *Glycerol phenylbutyrate* exhibited a notably strong predicted interaction with the AdoMet pocket, achieving a significantly more negative Vinardo score than molecules binding to the other sites.

3.2 Results of EvoFLOPA and Novel Inhibitor Identification

The EvoFLOPA algorithm was applied to optimize lead molecules for not only for binding to the AdoMet pocket of TrmD but for drug-like properties as well. The algorithm was run for a little over 500 iterations in the first run and for 1000 iterations in the subsequent run.

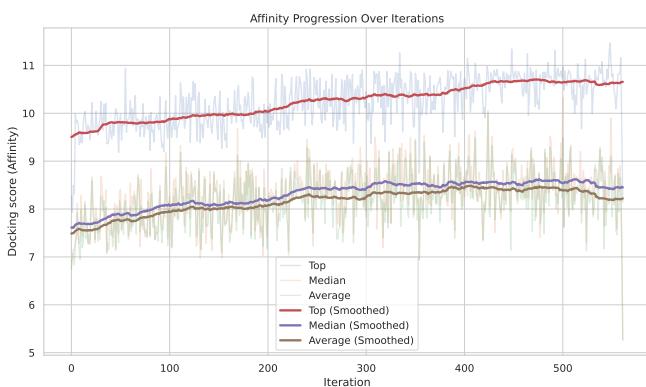


Figure 7: Binding affinity progression over 500 iterations for EvoFLOPA run no. 1.

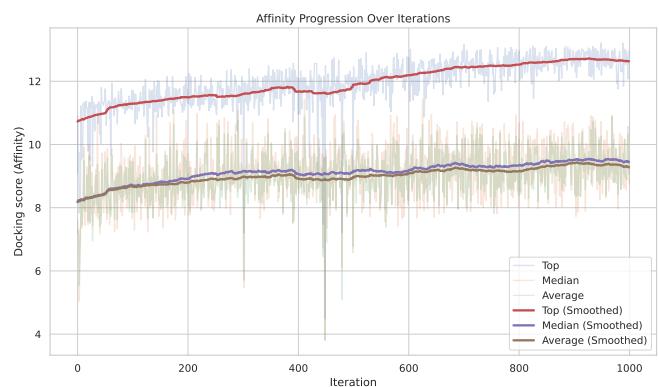


Figure 8: Binding affinity progression over 1000 iterations for EvoFLOPA run no. 2.

Figures 7 and 8 illustrate the progression of docking scores over two EvoFLOPA runs. The graphs demonstrate a trend of improving docking scores as the algorithm progressed, indicating successful optimization of binding affinity.

EvoFLOPA identified a novel compound, designated **Compound YF**, which exhibited a significantly improved RTMScore of 42.767, surpassing all FDA-approved molecules and known literature compounds, including Compound 23 from Wilkinson et al. (RTMScore: 32.792). Compound YF achieved a UniDock score of -12.72 kcal/mol, synthetic accessibility score of 3.05, and a QED score of 0.603.

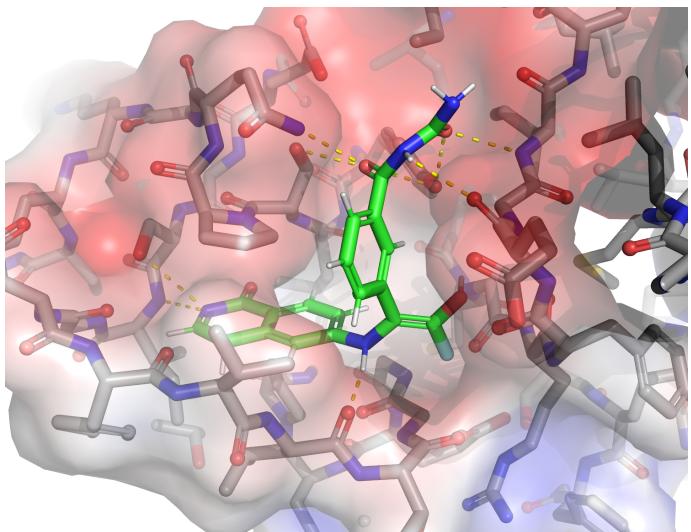


Figure 9: Docked Compound YF and its interactions with the AdoMet binding site.

Figure 9 displays the predicted pose and key interactions of Compound YF inside the AdoMet binding pocket. However, DILI prediction revealed a potential concern regarding bile salt export pump inhibition, suggesting potential hepatotoxicity.

Further optimization efforts focused on modifying Compound **YF** to mitigate the DILI risk while maintaining high binding affinity. This led to the design of Compound **YFOH**, which incorporates an additional hydroxyl (-OH) group.

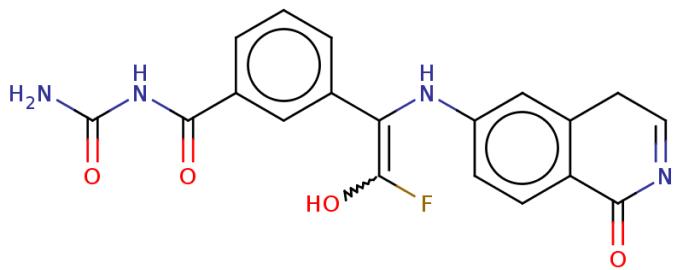


Figure 10: Chemical structures of Compound YFOH

While UniDock failed to dock **Compound YFOH** into the TrmD AdoMet pocket effectively, docking with SwissDock, a different docking program, successfully predicted a binding pose. Subsequent redocking of YFOH using UniDock yielded a score of -10.922 kcal/mol, and rescoring with RTMScore resulted in a score of 42.04.

Table 2: Comparison of Top Ligands Based on Various Scores

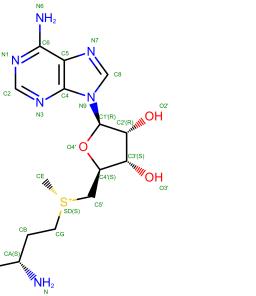
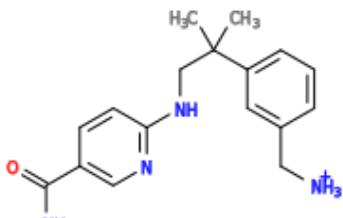
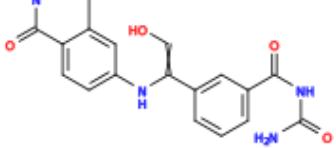
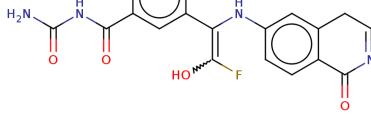
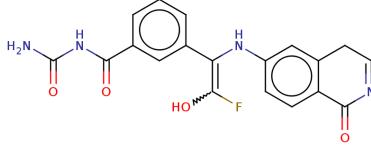
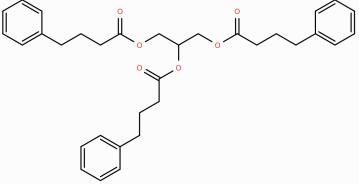
Ligand Name	Structure	Vina Score	QED	SA Score	RTMScore
Adomet		-8.69	0.29	4.74	27.6
Compound 23		-8.61	0.76	2.30	34.57
Compound Y		-13.10	0.62	3.01	38.6
Compound YF		-12.72	0.60	3.06	42.8
Compound YOH		-13.06	0.51	3.06	40.2
Ravicti		-13.50	0.16	3.16	32.2

Table 2. provides a comparative overview of the key properties of the identified novel compounds alongside top FDA-approved molecules and Compound 23 from Wilkinson et al., highlighting the superior RTMScore achieved by EvoFLOPA-generated compounds.

4 Discussion

This study successfully demonstrated the application of a computational pipeline, incorporating virtual screening and the EvoFLOPA evolutionary algorithm, for the discovery of novel TrmD inhibitors.

The benchmarking phase using FDA-approved molecules validated our docking protocol and provided a reference point for evaluating the performance of EvoFLOPA. The histogram analysis of docking scores revealed distinct binding preferences across the three TrmD sites, with the AdoMet pocket showing the most promise for accommodating diverse ligands.

EvoFLOPA proved to be an effective tool for lead optimization, successfully navigating chemical space and identifying novel compounds with enhanced predicted binding affinity. Compound YF, identified by EvoFLOPA, achieved a significantly higher RTMScore than both FDA-approved molecules and previously reported TrmD inhibitors, indicating a potentially superior binding interaction with the TrmD target. The use of SELFIES representation in EvoFLOPA ensured the generation of chemically valid molecules throughout the optimization process, a crucial aspect for practical drug discovery.

The validation of docking poses using RTMScore further strengthened our findings. RTMScore, as a deep learning-based scoring function, provides a more refined and potentially more accurate assessment of binding affinity compared to empirical scoring functions like Vinardo. The observation that RTMScore favored specific structural modifications, such as fluorine and hydroxyl group additions, offers valuable insights for further lead optimization.

While Compound YF exhibits promising predicted binding affinity and synthetic accessibility, the DILI prediction highlights a potential liability. The subsequent optimization to Compound YFOH, while requiring a different docking approach (SwissDock) for initial pose generation, retained high predicted affinity and may offer improved ADMET properties, warranting further investigation.

It is important to acknowledge the inherent limitations of in silico studies. Docking scores and ADMET predictions are estimations and require experimental validation. The stochastic nature of evolutionary algorithms also means that results may vary across different runs. Nevertheless, EvoFLOPA provides a powerful platform for prioritizing and accelerating the identification of promising lead candidates for experimental validation.

Using TrmD as an antibacterial target has some challenges. Although TrmD is essential for bacterial growth and lacks a human homolog, some studies show that bacteria can adapt to TrmD deficiency. For example, experimental evolution has shown that *E. coli* can survive without TrmD activity by compensatory mutations in other genes like proS, which restores tRNA function.[16] This ability of bacteria to evolve resistance mechanisms underscores the importance of exploring strategies that can overcome these limitations. Therefore, we suggest that a promising approach could be a combined therapy that targets both TrmD and the compensatory pathway involving ProS. Inhibiting both TrmD and ProS concurrently would likely hinder the bacteria's ability to adapt and maintain tRNA function, potentially leading to a more effective and durable antibacterial treatment. By simultaneously targeting both the primary target and a crucial compensatory mechanism.

5 Conclusion

This study demonstrates the successful development and application of EvoFLOPA, an evolutionary algorithm-based pipeline, for the computational discovery of novel TrmD inhibitors. EvoFLOPA effectively optimized lead molecules, leading to the identification of Compounds YF and YFOH, which exhibit significantly improved predicted binding affinity to TrmD compared to FDA-approved molecules and previously known inhibitors. Rescoring with RTMScore further validated these findings, highlighting the potential of these novel compounds as promising leads for antibacterial drug development.

While TrmD is an essential target for bacterial growth and lacks a human homolog, its inhibition alone may not be sufficient due to the ability of bacteria to evolve compensatory mechanisms, such as mutations in the ProS pathway that restore tRNA function. To address this limitation, a dual-targeting strategy is suggested, where both TrmD and the compensatory pathway involving ProS are inhibited concurrently. This approach could effectively hinder bacterial adaptation, enhancing the durability and efficacy of antibacterial therapies.

While further experimental validation is crucial, this work not only highlights the potential of the identified compounds but also underscores the importance of considering compensatory pathways in target selection and drug development. The integration of evolutionary algorithms and computational approaches represents a powerful strategy for accelerating the discovery of novel therapeutics to combat antibiotic resistance.

6 Acknowledgements

We would like to thank our supervisor, Mgr. Marian Novotný, Ph.D., for his guidance and support throughout the TrmD project. His expertise and valuable insights have contributed to the progress and completion of this work. Last but not least, special thanks to RNDr. Mgr. Martin Lepšík, Ph.D. for his helpful advice during the docking phase.

References

- [1] World Health Organization. Antimicrobial Resistance Fact Sheet, 2024. URL <https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance>. Accessed: 2024-01-25.
- [2] Hyeon Jeong Ahn, Hyeon Seok Kim, Hyun Kyu Yoon, and Byung Il Lee. Structural basis for the catalytic mechanism of trna (m1g37) methyltransferase (trmd). *Acta Crystallographica Section D: Structural Biology*, 75(Pt 8):747–758, 2019. doi: 10.1107/S2059798319009940. URL <https://doi.org/10.1107/S2059798319009940>.
- [3] Shigeki Goto-Ito, Takashi Ito, Ryo Ishii, Yugo Muto, Yuko Bessho, and Shigeyuki Yokoyama. Trm5 and trmd: Two enzymes from distinct origins catalyze the identical trna modification. *Nucleic Acids Research*, 45(8):4739–4748, 2016. doi: 10.1093/nar/gkw1180. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5372744/>.
- [4] Huifeng Zhao, Dejun Jiang, Chao Shen, Jintu Zhang, Xujun Zhang, Xiaorui Wang, Dou Nie, Tingjun Hou, and Yu Kang. Comprehensive evaluation of 10 docking programs on a diverse set of protein–cyclic peptide complexes. *Journal of Chemical Information and Modeling*, 64(6):2112–2124, 2024.
- [5] S. Kim, J. Lee, and H. Nam. Leadd: A lamarckian evolutionary algorithm for de novo molecular design. *Journal of Cheminformatics*, 14(1):82, 2022. doi: 10.1186/s13321-022-00582-y. URL <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-022-00582-y>.
- [6] Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alán Aspuru-Guzik. SELFIES: A robust representation of semantically constrained graphs with an example application in chemistry. *Machine Learning: Science and Technology*, 1(4):045024, 2020. doi: 10.1088/2632-2153/aba947. URL <https://doi.org/10.1088/2632-2153/aba947>.
- [7] T. Ito and S. Yokoyama. Crystal structure of *H. influenzae* trmd in complex with adomet. *Proceedings of the National Academy of Sciences*, 112(41):E5565–E5573, 2015. doi: 10.1073/pnas.1422981112. URL <https://doi.org/10.1073/pnas.1422981112>. PDB DOI: <https://doi.org/10.2210/pdb4yvg/pdb>.
- [8] Radoslav Krivák and David Hoksza. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *Journal of Cheminformatics*, 10(1):39, 2018. doi: 10.1186/s13321-018-0285-8. URL <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0285-8>.
- [9] Takashi Ito, Isao Masuda, Kenji Yoshida, Shigeki Goto-Ito, Shuichi Sekine, Seong-Jin Suh, Ya-Ming Hou, and Shigeyuki Yokoyama. Structural basis for methyl-donor–dependent and sequence-specific recognition of substrate trna by trna (m1g37) methyltransferase (trmd). *Proceedings of the National Academy of Sciences*, 112(27):E4197–E4205, 2015. doi: 10.1073/pnas.1422981112. URL <https://doi.org/10.1073/pnas.1422981112>.
- [10] RDKit: Open-source cheminformatics. <https://www.rdkit.org>. Accessed: 2025-01-25.
- [11] Yuejiang Yu, Chun Cai, Jun Wang, Zhi Bo, Zhengdan Zhu, and Hang Zheng. Uni-dock: Gpu-accelerated docking enables ultralarge virtual screening. *Journal of Chemical Theory and Computation*, 19(11):3336–3345, 2023. doi: 10.1021/acs.jctc.2c01145. URL <https://doi.org/10.1021/acs.jctc.2c01145>.
- [12] Benjamin Wilkinson et al. Evaluating the druggability of trmd, a potential antibacterial target, through design and microbiological profiling of a series of potent trmd inhibitors. *Bioorganic Medicinal Chemistry Letters*, 78:129331, 2023. doi: 10.1016/j.bmcl.2023.129331. URL <https://doi.org/10.1016/j.bmcl.2023.129331>.
- [13] AkshatKumar Nigam, Robert Pollice, Mario Krenn, Gabriel dos Passos Gomes, and Alán Aspuru-Guzik. Beyond generative models: Superfast traversal, optimization, novelty, exploration, and discovery (stoned) algorithm for molecules using selfies. *Chemical Science*, 12(20):7079–7090, 2021. doi: 10.1039/D1SC00231G. URL <https://doi.org/10.1039/D1SC00231G>.
- [14] Chao Shen, Xujun Zhang, Yafeng Deng, Junbo Gao, Dong Wang, Lei Xu, Peichen Pan, Tingjun Hou, and Yu Kang. Boosting protein–ligand binding pose prediction and virtual screening based on residue–atom distance likelihood potential and graph transformer. *Journal of Medicinal Chemistry*, 65(15):10691–10706, 2022.
- [15] Guangxu Xiong, Zhiwei Wu, Junfeng Yi, Lianding Fu, Zhen Yang, Chia-Hsiu Hsieh, Mingzhu Yin, Xinkai Zeng, Chenguang Wu, Aiping Lu, Xijun Chen, Tingjun Hou, and Dong-Sheng Cao. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research*, 49(W1):W5–W14, 2021. doi: 10.1093/nar/gkab255. URL <https://doi.org/10.1093/nar/gkab255>.
- [16] Ben E. Clifton, Ben E. Clifton, Muhammad Aiman Fariz, Muhammad Aiman Fariz, Gen ichiro Uechi, Gen-Ichiro Uechi, Paola Laurino, and Paola Laurino. Evolutionary repair reveals an unexpected role of the trna modification m1g37 in aminoacylation. *Nucleic Acids Research*, 2021. doi: 10.1093/nar/gkab1067.