Meet-EU 2024 Final Report

Szymon Jakubicz, Przemysław Pietrzak,
Daniel Zalewski, Katsiaryna Dubrouskaya

Abstract

This study explores computational approaches to discover novel inhibitors of the bacterial enzyme TrmD, a target for combating antibiotic resistance. Starting with 2.43 million compounds from the ChEMBL database, iterative filtering and docking methods, including DiffDock-L and STONED, reduced the dataset to 5,000 validated molecules. ADMET analysis shortlisted 10 candidates, with molecule C demonstrating favorable pharmacological properties. LigPlot analysis revealed strong binding interactions with TrmD, and comparison with its parent molecule showed enhanced stability due to additional hydrophobic bonds. This work highlights the potential of computational tools in identifying selective and effective TrmD inhibitors.

1 Introduction

The global rise of bacterial antimicrobial resistance (AMR) poses a critical threat to public health. It is estimated that antibiotic-resistant bacterial infections caused 1.27 million deaths worldwide in 2019 and contributed to 4.95 million total deaths [20]. The Review on Antimicrobial Resistance, commissioned by the UK Government, argued that AMR could kill 10 million people per year by 2050 [23]. To address this crisis, researchers are exploring novel targets for antibiotic development. One promising area is post-transcriptional RNA modifications, particularly tRNA modifications. TrmD, a tRNA methyltransferase enzyme, which is essential for growth and highly conserved in both Gram-positive and Gram-negative bacterial pathogens [12], [9], [27], stands out as an attractive target.

TrmD is a tRNA methyltransferase enzyme that is responsible for methylating guanine at position 37 to form 1-methylguanosine (m1G37) in tRNAs containing a G36G37 motif, using S-adenosyl-L-methionine (SAM) as a methyl donor [3]. The methylated m1G37 is on the 3-side of the anticodon, and it is necessary for suppressing tRNA frameshifting during protein synthesis on the ribosome [10]. TrmD possesses a unique deep trefoil knot structure in its N-terminal domain, which is essential for binding S-adenosylmethionine (SAM). Unlike other methyltransferases, TrmD requires magnesium ions ($Mg2+$) in its catalytic mechanism [24]. The enzyme forms an obligate homodimer, with active sites located at the dimer interface. Its structure and mechanism are fundamentally distinct from its eukaryotic counterpart, Trm5, despite catalyzing the same chemical reaction [14].

Machine learning approaches are increasingly influential in drug discovery processes, particularly for in silico methods [7]. These techniques include:

1. Artificial neural networks, support vector machines, and deep learning for predicting pharmacological properties and molecular docking outcomes.

2. Potential to analyze large libraries of diverse chemical structures and handle complex, high-dimensional data.

3. Applications in retrosynthesis planning and de novo drug design.

However, challenges remain, including data quality issues and the interpretability of complex models. Despite these limitations, machine learning is increasingly central in computational drug discovery, complementing traditional structure-based and quantitative structure–activity relationship (QSAR) methodologies [19].

Together, the unique attributes of TrmD and the ML-driven computational drug discovery offer approaches to address antibiotic resistance through selective and efficient drug design.

2 Methods

2.1 Initial screening

The initial screening was conducted using the ChEMBL database version 34, which comprises 2.431.025 molecules [11]. To identify compounds with favorable drug-like properties, a series of

stringent filters were applied based on established pharmacokinetic and chemical criteria.

First, Lipinski's Rule of Five was employed to ensure good oral bioavailability by selecting molecules with a molecular weight of 500 Dalton or less, no more than five hydrogen bond donors, no more than ten hydrogen bond acceptors, and a logarithm of the partition coefficient ($\log P$) of five or lower [18]. This rule serves as a fundamental guideline in drug discovery to predict the drug-likeness of compounds.

Subsequently, the Ghose Filter was utilized to further refine the selection by retaining compounds with a molecular weight between 160 and 480 Dalton, logP values ranging from -0.4 to 5.6, a total number of atoms between 20 and 70, and a polar surface area between 60 and 130 $Å^2$ [13]. This filter helps in identifying molecules that fall within the typical property ranges of known drug-like substances.

To eliminate reactive, unstable, or otherwise undesirable molecules, the REOS Filter was applied, which excludes compounds containing reactive functional groups or structural alerts associated with toxicity or instability [29]. This step is crucial for enhancing the safety profile of the selected compounds.

Enhancing the potential for oral bioavailability, the Veber Filter was implemented, selecting molecules with ten or fewer rotatable bonds and a polar surface area of 140 $Å^2$ or less [30]. The Veber criteria are effective in predicting the oral bioavailability of drug candidates by focusing on molecular flexibility and polarity.

The Drug-like Filter was then used to ensure that the remaining compounds possess a combination of physicochemical properties typical of known drugs, thereby increasing the likelihood of therapeutic efficacy [33]. This general filter consolidates various drug-like attributes to streamline the selection process.

To quantitatively assess the overall drug-likeness of the compounds, the Quantitative Estimate of Drug-likeness (QED) filter was applied, retaining molecules with a QED score of 0.7 or higher [1]. The QED score integrates multiple molecular properties to provide a comprehensive measure of drug-likeness.

Additionally, the Synthetic Accessibility (SA) Score Filter was employed to evaluate the ease of synthesizing the compounds, selecting those with an SA score of three or lower to ensure reasonable synthetic feasibility [8]. This consideration is vital for practical drug development, where the ease of synthesis can significantly impact the feasibility and cost of producing the compounds.

Finally, the Natural Product-likeness (NP) Score Filter was used to identify molecules resembling natural products, which often exhibit favorable biological activities, by including compounds with an NP score of 0.2 or higher [4]. Natural product-like compounds are valuable in drug discovery due to their structural diversity and biological relevance.

Through the application of these comprehensive filters, the dataset was reduced to 8.5k molecules. The entire filtering process was executed on a high-performance computing cluster, ensuring efficiency and scalability. Importantly, the filtering criteria were applied in an agnostic manner, independent of any specific biological targets or known inhibitors. This approach ensured an unbiased selection of compounds based solely on their intrinsic chemical and pharmacokinetic properties, facilitating the identification of novel candidates with potential therapeutic value.

2.2 Molecular docking using Machine Learning model

Molecular docking is a technique used to predict the orientation of a molecule (ligand) when it binds to another protein. We performed molecular docking three times at different stages of the project. In the first experiment we sorted the ligands from Initial Screening using their Quantitative Estimate of Druglikeness [2] obtaining 8.5k ligands taken from in the SMILES format. We also used a .pbd file with bacterial protein. In the second experiment, we used 1.5k ligands obtained from STONED and a .pdb file with a human protein. Last time we used eighteen inhibitors as ligands with the bacteria protein. Preparing data needed processing of the input to a file that has a structure that the method expects.

Experiments were carried out using DiffDock-L [5] method — improved DiffDock[6] method. DiffDock models takes separate ligand and protein structure and denoises randomly sampled initial poses by reverse diffusion over translations, rotations and torsions. Then the poses are ranked by the confidence model to produce a final prediction and confidence score. The confidence score is returned from confidence model used by DiffDock. There is Spearman correlation of 0.68 between confidence and negative RMSD.

We cloned the code from the DiffDock repository. We were unable to run DiffDock on faculty's

Entropy server because running docker code on singularity was very unhandy and we don't have root privileges there to install it normally so experiments were performed on our local GPUs. On one computer with a good enough GPU to run DiffDock in a reasonable time we had to install DiffDock in the Windows Subsystem for Linux environment so some additional configuration had to be performed like installing multiple dependencies manually. After the DiffDock configuration ran it was easy as it was performed using the command provided in the repository. We used the default configuration file, as increasing the prediction precision led to a long execution time while decreasing it led to the model being unable to experiment successfully.

2.3 Molecular Generation and Validation

Following the initial docking process with DiffDock, only molecules exhibiting a confidence score greater than zero were retained, resulting in a base set of 83 molecules. These high-confidence molecules served as the foundation for subsequent molecular generation and optimization steps.

To expand the chemical space around these base molecules, the STONED (String-based Techniques for Optimization and Exploration of Novel Drug-like Entities) method was employed [22]. The STONED method leverages the robustness of the SELFIES (Self-referencing Embedded Strings) [21] representation to facilitate the perturbation and generation of novel molecular structures. By systematically modifying the SELFIES strings, STONED ensures that the generated molecules maintain chemically valid structures while exploring diverse chemical modifications.

Utilizing STONED, a total of 388.440 molecules were generated, of which 281.634 were unique. This substantial expansion was achieved by creating eight new molecules from each molecule in the previous mutation stage, thereby exponentially increasing the diversity of the chemical library.

The generated molecules underwent a stringent filtering process to ensure chemical diversity and relevance. Specifically, molecules were filtered based on Tanimoto similarity using selected fingerprinting methods, namely Extended-Connectivity Fingerprints (ECFP6) [26] and PATH fingerprints [17]. The filtering criteria were defined as follows: for ECFP6 fingerprints, Tanimoto similarity scores between 0.5 and 0.8; for PATH fingerprints, Tanimoto similarity scores between 0.5 and 0.8. This dual-filtering approach ensured that the sampled molecules were neither too similar nor too dissimilar to the base molecules, maintaining a balanced exploration of the chemical space [26][17]
.

Subsequent to similarity filtering, the SYBA (Synthetic Balanced Accuracy) score model was applied to prioritize the 7.627 molecules with the most favorable synthetic accessibility and drug-like properties [31] obtained in the previous stage. The SYBA model integrates various molecular descriptors to assess the feasibility of synthesizing the compounds, thereby streamlining the selection of promising candidates for further evaluation.

As part of the molecule validation process, an attempt was made to convert the selected molecules into their three-dimensional (3D) structures. This conversion was performed using the RDKit package [16], which facilitates the generation and manipulation of molecular structures in the SDF (Structure Data File) format. The validation process was iteratively conducted until 5k valid molecules were obtained, at which point the generation and validation pipeline was halted. These validated molecules were subsequently selected for in-depth examination and potential experimental testing.

2.4 ADMET analysis and comparison with a known molecule

For the ADMET [15] analysis, 503 molecules were selected that achieved a confidence score > 0.0 for TrmD and a confidence score < 0.0 for Trm5. The molecules were further analyzed based on solubility, permeability, absorption, toxicity, and interactions with CYP enzymes. We utilized the pkCSM [25] tool available at `https://biosig.lab.uq.edu.au/pkcsm/`. The molecules underwent pkCSM filters covering water solubility, Caco-2 permeability [25], intestinal absorption, AMES toxicity, hepatotoxicity, LD50, VDss, and substrate or inhibitor properties for CYP [25] enzymes. Ten molecules passed all the pkCSM filters, and the molecule with the best docking score was selected for structural analysis. We used LigPlot [32] for the analysis of protein-ligand interactions. Additionally, we performed a structural analysis relative to the known parent molecule in Pymol [28].

3 Results

3.1 ADMET

The ADMET results are presented in Table 3.1. The analyzed molecules met all selection criteria. The molecules exhibited good water solubility (> -4 log mol/L) and Caco-2 permeability ($> 1 \log$ Papp in $10^{-6} \frac{cm}{s}$), indicating favorable membrane permeability. Furthermore, the evaluation of intestinal absorption ($> 80\%$) and VDss ($0 \log \frac{L}{kg} <$ VDss $< 2 \log \frac{L}{kg}$) excluded compounds with poor oral bioavailability. The molecules demonstrated low AMES toxicity, hepatotoxicity and high LD50 value ($> 2 \frac{mol}{kg}$), which enhances the safety profile of the compounds. The molecules are neither substrates nor inhibitors of CYP enzymes. The structural formulas of the generated molecules are presented in Figure 1. Among them, molecule C was selected for further analysis.

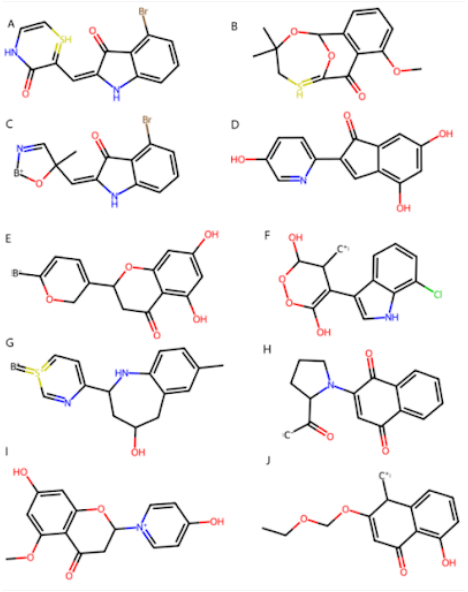

Figure 1: Structural formulas of the molecules identified after ADMET filtering.

SMILES	Water solubility	Caco2 permeability	Intestina absorption	VDss	CYP substrate	CYP inhibitor	AMES	LD50	Hepatotoxicity
O=C1NC=C[SH]=C1C=C1Nc2cccc(Br)c2C1=O	-3.613	1.304	92.963	0.063	No	No	No	2.373	No
COc1cccc2c1C(=O)C1=[SH]CC(C)(C)OC2O1	-3.373	1.308	93.812	0.034	No	No	No	2.331	No
CC1(C=C2Nc3cccc(Br)c3C2=O)C=N[B+]O1	-3.931	1.318	93.655	0.032	No	No	No	2.842	No
O=C1C(c2ccc(O)cn2)=Cc2c(O)cc(O)cc21	-2.674	1.034	91.804	0.014	No	No	No	2.269	No
[B-]C1=CC=C(C2CC(=O)c3c(O)cc(O)cc3O2)CO1	-2.367	1.101	93.955	0.056	No	No	No	2.262	No
[C+]C1C(c2c[nH]c3c(Cl)cccc23)=C(O)OOC1O	-3.77	1.024	92.261	0.412	No	No	No	2.094	No
[B+]=[S+]1=CC=C(C2CC(O)Cc3cc(C)ccc3N2)N=C1	-3.592	1.232	95.292	0.503	No	No	No	2.469	No
[C-]C(=O)C1CCCN1C1=CC(=O)c2ccccc2C1=O	-2.617	1.261	96.57	0.085	No	No	No	2.351	No
COc1cc(O)cc2c1C(=O)CC([n+]1ccc(O)cc1)O2	-2.296	1.191	93.112	0.779	No	No	No	3.16	No
[C+]C1C(OCOCC)=CC(=O)c2c(O)cccc21	-2.538	1.261	95.644	0.04	No	No	No	2.046	No

Table 1: The top 10 molecules from ADMET analysis. The indicated molecules exhibit the best distribution, absorption and toxicity properties.

3.2 Protein-ligand interactions

The next step was the LigPlot analysis to determine the interactions between the newly generated molecule C and TrmD. The LigPlot results are presented in Figure 2. The key hydrogen bonds involved interactions of the ligand with Leu 138(A), specifically the ligand's nitrogen atom interacting with the carbonyl and amino groups of Leu 138(A), with bond lengths of 3.3Å and 3Å. These bonds stabilize the ligand within the binding site. The main amino acid interacting hydrophobically with the ligand was Tyr 115(A). Due to the proximity of Tyr 115(A), which contains a phenolic ring, to the highly polarizable bromine atom, strong van der Waals interactions are observed. Additionally, due to the proximity of Gly 141(A) to the ligand, this amino acid also contributes to stabilization.

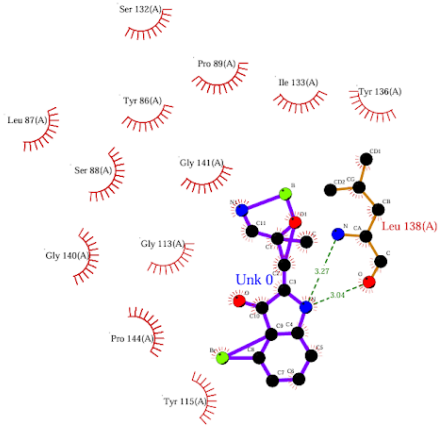

Figure 2: LigPlot showing hydrophobic and hydrogen bonds between the TrmD and molecule C.

3.3 Comparison of the molecule with a known compound

The parent molecule for molecule C was CHEMBL3798674. A comparison of both molecules is presented in Figure 3. CHEMBL3798674 is one of the inhibitors of Serine/threonine-protein kinase 17B, with bonds (1,2). The similarity between the molecules is approximately 80%. Both molecules utilize the nitrogen atom (1) to form hydrogen bonds with TrmD and the polarizable bromine atom (2) to form van der Waals interactions. An advantage of the new molecule is the formation of additional hydrophobic interactions at the nitrogen atom (3). As a result of this additional bond, the newly generated molecule C demonstrates enhanced stability in binding to the active site of TrmD.

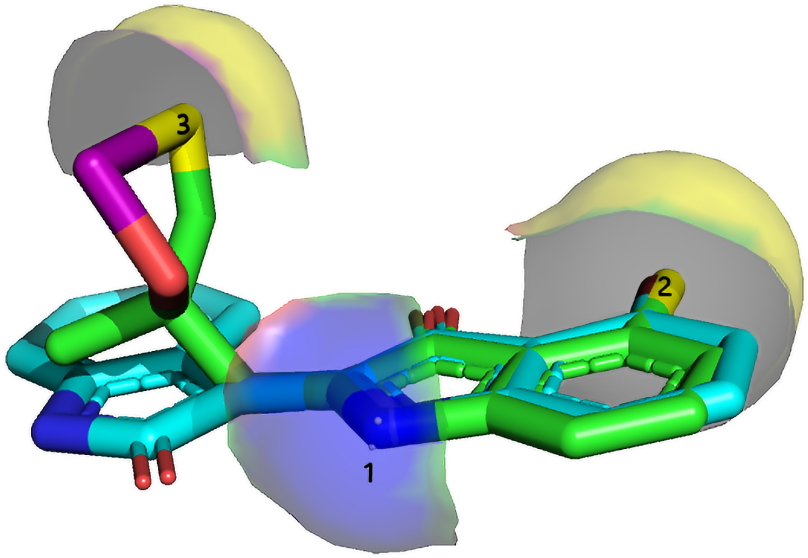

Figure 3: Comparison of molecule C and CHEMBL3798674 interactions with the TrmD protein.

4 Conclusion

The search for new inhibitors and a potential drug for drug-resistant bacterial strains is ongoing. Our approach utilized DiffDock and STONED as representatives of innovative ML applications in drug design. Controlled generation of new molecules using STONED enables the analysis of many previously untested compounds to identify the best bindings to TrmD. The molecules proposed by our team demonstrate favorable ADMET properties and progress in creating new bonds with the molecular target.

Our research could be further developed by conducting additional iterations of STONED and DiffDock, selecting the best parent molecule for generating new compounds. Moreover, our molecules have not yet undergone molecular dynamics simulation analysis, which would serve as a valuable benchmark before conducting in vitro studies. Each iteration could be enhanced with a screening process using filters that consider molecular structure, thereby increasing the generative capabilities of the pipeline.

5 Authors' Contributions

- Daniel Zalewski: Team management, project consultation and communication with course coordinators, study design and concept, ADMET analysis and final results interpretation, contribution to writing the report, preparation of the presentation and flash talk for the conference.
- Przemysław Pietrzak: Review of ML methods and suggestion of the docking model used in the in silico screening pipeline, contribution to data and code management, initial screening with established metrics, generation of derivative molecules based on mutation of the filtered molecules, contribution to writing the final report.

- Szymon Jakubicz: Contribution in review on ML methods, contribution to data and code management, conducting docking model experiments on private device, review and transfer to correct format of the final report, creating poster
- Katsiaryna Dubrouskaya: Conducting docking model experiments on private device, contribution to data management, contribution to writing the report, creating poster

References

[1] G. R. Bickerton, G. V. Paolini, D. Williams, C. R. Ashby, C. J. Williams, and G. R. Williams. Qed: a multi-property scoring function for drug-likeness based on a detailed analysis of 2000 real drugs. *Journal of Medicinal Chemistry*, 55(4):3649–3668, 2012. doi:10.1021/jm2010572.

[2] Geoffrey R. Bickerton, Giovanni V. Paolini, Julien Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012. doi:10.1038/nchem.1243.

[3] Anders S Byström and Glenn R Björk. The structural gene (trmd) for the trna (m1g) methyltransferase is part of a four polypeptide operon in escherichia coli k-12. *Molecular and General Genetics MGG*, 188(3):447–454, 1982.

[4] V. Consonni and R. Todeschini. Natural product likeness score: A single parameter descriptor for natural product-like compounds. *Journal of Chemical Information and Modeling*, 50(10):1778–1786, 2010. doi:10.1021/ci100143f.

[5] Gabriele Corso, Arthur Deng, Benjamin Fry, Nicholas Polizzi, Regina Barzilay, and Tommi Jaakkola. Deep confident steps to new pockets: Strategies for docking generalization, 2024. URL: https://arxiv.org/abs/2402.18396, arXiv:2402.18396.

[6] Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking, 2023. URL: https://arxiv.org/abs/2210.01776, arXiv:2210.01776.

[7] Jacob D Durrant and Rommie E Amaro. Machine-learning techniques applied to antibacterial drug discovery. *Chemical biology & drug design*, 85(1):14–21, 2015.

[8] P. Ertl and A. Schuffenhauer. Estimation of synthetic accessibility of drug-like molecules. *European Journal of Medicinal Chemistry*, 44(1):163–169, 2009. doi:10.1016/j.ejmech.2008.07.046.

[9] R Allyn Forsyth, Robert J Haselbeck, Kari L Ohlsen, Robert T Yamamoto, Howard Xu, John D Trawick, Daniel Wall, Liangsu Wang, Vickie Brown-Driver, Jamie M Froelich, et al. A genome-wide strategy for the identification of essential genes in staphylococcus aureus. *Molecular microbiology*, 43(6):1387–1400, 2002.

[10] Howard B Gamper, Isao Masuda, Milana Frenkel-Morgenstern, and Ya-Ming Hou. Maintenance of protein synthesis reading frame by ef-p and m1g37-trna. *Nature communications*, 6(1):7226, 2015.

[11] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, and J. P. Overington. Chembl: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 45(D1):D945–D954, 2017. doi:10.1093/nar/gkw1039.

[12] SYea Gerdes, MD Scholle, JW Campbell, G Balazsi, E Ravasz, MD Daugherty, AL Somera, NC Kyrpides, I Anderson, MS Gelfand, et al. Experimental determination and system level analysis of essential genes in escherichia coli mg1655. *Journal of Bacteriology*, 185(19):5673–5684, 2003. doi:10.1128/jb.185.19.5673-5684.2003.

[13] A. Ghose, V. N. Viswanadhan, and S. Engel. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. *Journal of Medicinal Chemistry*, 42(24):3118–3131, 1999. doi:10.1021/jm990254j.

[14] Sakurako Goto-Ito, Takuhiro Ito, and Shigeyuki Yokoyama. Trm5 and trmd: two enzymes from distinct origins catalyze the identical trna modification, m1g37. *Biomolecules*, 7(1):32, 2017.

[15] L Guan, H Yang, Y Cai, L Sun, P Di, W Li, G Liu, and Y Tang. Admet-score - a comprehensive scoring function for evaluation of chemical drug-likeness. *MedChemComm*, 10(1):148–157, 2018. doi:10.1039/c8md00472b.

[16] G. Landrum. Rdkit: Open-source cheminformatics. *Journal of Cheminformatics*, 8(1):1, 2016. doi:10.1186/s13321-016-0169-8.

[17] Greg Landrum. Rdkit: Open-source cheminformatics. *Journal of Cheminformatics*, 8(1):1, 2016. doi:10.1186/s13321-016-0169-8.

[18] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23:3–25, 1997.

[19] Yu-Chen Lo, Stefano E. Rensi, Wen Torng, and Russ B. Altman. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today*, 23(8):1538–1546, 2018. URL: https://www.sciencedirect.com/science/article/pii/S1359644617304695, doi:10.1016/j.drudis.2018.05.010.

[20] Christopher JL Murray, Kevin Shunji Ikuta, Fablina Sharara, Lucien Swetschinski, Gisela Robles Aguilar, Authia Gray, Chieh Han, Catherine Bisignano, Puja Rao, Eve Wool, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The lancet*, 399(10325):629–655, 2022.

[21] A. Nigam, R. Pollice, M. Krenn, G. Dos Passos Gomes, and A. Aspuru-Guzik. Beyond generative models: superfast traversal, optimization, novelty, exploration and discovery (stoned) algorithm for molecules using selfies. *Chemical Science*, 14:3446–3459, 2023.

[22] 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. *Chem. Sci.*, 12:7079–7090, 2021. URL: http://dx.doi.org/10.1039/D1SC00231G, doi:10.1039/D1SC00231G.

[23] Jim O'Neill. Tackling drug-resistant infections globally: final report and recommendations. 2016.

[24] Agata P Perlinska, Marcin Kalek, Thomas Christian, Ya-Ming Hou, and Joanna I Sulkowska. Mg2+-dependent methyl transfer by a knotted protein: A molecular dynamics simulation and quantum mechanics study. *ACS catalysis*, 10(15):8058–8068, 2020.

[25] Douglas E. V. Pires, Tom L. Blundell, and David B. Ascher. pkcsm: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *Journal of Medicinal Chemistry*, 58(9):4066–4072, 2015. PMID: 25860834. arXiv:https://doi.org/10.1021/acs.jmedchem.5b00104, doi:10.1021/acs.jmedchem.5b00104.

[26] David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010.

[27] Christopher M Sassetti, Dana H Boyd, and Eric J Rubin. Genes required for mycobacterial growth defined by high density mutagenesis. *Molecular microbiology*, 48(1):77–84, 2003.

[28] Schrödinger, LLC. The PyMOL molecular graphics system, version 1.8. November 2015.

[29] N. Vargesson. Reos: a filtering tool to remove compounds with undesirable reactivity. *ChemMedChem*, 4(5):757–760, 2009. doi:10.1002/cmdc.200900027.

[30] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of Medicinal Chemistry*, 45(12):2615–2623, 2002. doi:10.1021/jm020051v.

[31] Petr Voronov, Lenka Duchonova, Rastislav Jurik, Radka Svobodova Varekova, and Stanislav Geidl. Syba: Bayesian estimation of synthetic accessibility of organic compounds. *Journal of Cheminformatics*, 11(1):1–13, 2019.

[32] A. C. Wallace, R. A. Laskowski, and J. M. Thornton. Ligplot: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering*, 8:127–134, 1996. PubMed id: 7630882.

[33] W. P. Walters, M. A. Murcko, and D. W. Wright. Chemical space navigated through two-dimensional molecular frameworks. *Journal of Chemical Information and Modeling*, 48(2):362–368, 2008. doi:10.1021/ci700369m.