A UNIFIED COMPUTATIONAL FRAMEWORK FOR THE INTEGRATION OF AI MODELS IN STRUCTURE-BASED DRUG DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

The rapid proliferation of sophisticated computational models for drug discovery has created unprecedented opportunities for innovation, yet the field lacks comprehensive frameworks to systematically integrate these diverse tools into coherent workflows. Although these models demonstrate remarkable individual capabilities, researchers are forced to navigate fragmented toolsets requiring extensive computational expertise, limiting the practical impact of these advances and creating an accessibility problem. In this paper, we present a comprehensive and modular computational drug discovery pipeline that provides the first systematic framework for integrating diverse state-of-the-art models into an accessible unified drug discovery workflow. The workflow is based on the integration of state-of-the-art generative and docking models, with a special focus on ensuring the synthetic accessibility and real world scenario plausibility of the proposed molecules.

INTRODUCTION

The pharmaceutical industry faces increasing pressure to accelerate drug discovery while reducing the substantial costs and time requirements associated with traditional approaches (Sertkaya et al., 2024; Dowden & Munro, 2019). Computational drug discovery has emerged as a transformative solution, offering the potential for faster, more cost-effective development cycles (Vamathevan et al., 2019; Alquraishi et al., 2025) with reduced reliance on extensive in vivo testing (Gangwal & Lavecchia, 2025; Lee et al., 2022; U.S. Food and Drug Administration, 2025b;a).

Recently available molecular generative models (Loeffler et al., 2024; Cretu et al., 2025; Shen et al., 2025), molecular property predictors (Jin et al., 2025; Qiao et al., 2025), and docking models (Corso et al., 2023; Passaro et al., 2025), demonstrate remarkable capabilities within their specific domains, but despite this wealth of computational resources drug discovery lacks comprehensive frameworks and accessible tools that enable researchers to systematically integrate these diverse models into coherent workflows.

The full potential of generative and docking models remains unrealized due to fragmented toolsets and the scarcity of systematic frameworks for their integration with downstream validation and optimization processes. Another critical challenge in generative drug design is synthetic accessibility: many generative models propose molecular designs that, while computationally promising, prove impossible or prohibitively expensive to synthesize (Gao & Coley, 2020). This limitation severely undermines the real-world applicability of such approaches and represents a significant bottleneck in translating computational predictions to therapeutic reality.

Aurora kinases are renowned and challenging targets that play essential roles in cancer progression (Glover et al., 1995; Marumoto et al., 2005; Dar et al., 2010; Katayama et al., 2003; Yan et al., 2016). Despite decades of research (Carmena et al., 2015), existing Aurora kinase inhibitors face significant limitations, including lack of isoform selectivity leading to off-target toxicity, emergence of resistance mechanisms through gatekeeper mutations and compensatory pathways, and narrow therapeutic windows due to essential mitotic functions (Jing & Chen, 2021; Bavetsias & Linardopoulos, 2015). The convergence of abundant structural data in the Protein Data Bank (PDB, (Berman et al., 2000)) and well-characterized structure-activity relationships (Ashraf et al., 2021) make Aurora kinases an ideal testbed for computational design approaches.

In response to these challenges, we present a comprehensive and modular computational drug discovery pipeline that directly addresses the fragmentation issue in the field by providing the first systematic framework for integrating diverse state-of-the-art models into unified drug discovery workflows. Our approach bridges the gap between the abundance of available computational tools and their practical application, combining generative artificial intelligence with rigorous *post hoc* filtering and evaluation frameworks. By implementing this integration through an accessible Snakemake workflow (Mölder et al., 2021), we democratize access to sophisticated computational drug discovery approaches while providing researchers with the comprehensive logic and tools necessary to construct their own modular pipelines. We demonstrate the capabilities of this framework through systematic exploration of the chemical space for Aurora kinase inhibitor design, showcasing how diverse computational models can be orchestrated to address real-world drug discovery challenges. Figure 1 shows an overview of the pipeline. The code to reproduce the experiments is available at the following anonymized repository: `https://anonymous.4open.science/r/drug-design-pipeline-BF4F`.

Our contributions include:

- The conceptualization and practical implementation of a modular and accessible drug design pipeline integrating multiple state-of-the-art computational methods;
- The development of a comprehensive *post hoc* filtering logic incorporating synthesizability, drug-likeness, and binding affinity assessments;
- The creation of an inclusive Snakemake pipeline that democratizes access to advanced computational drug discovery tools;
- The systematic evaluation of generative model performance in real-world drug discovery scenarios, providing insights into the practical applicability of current AI-driven approaches to therapeutic development.

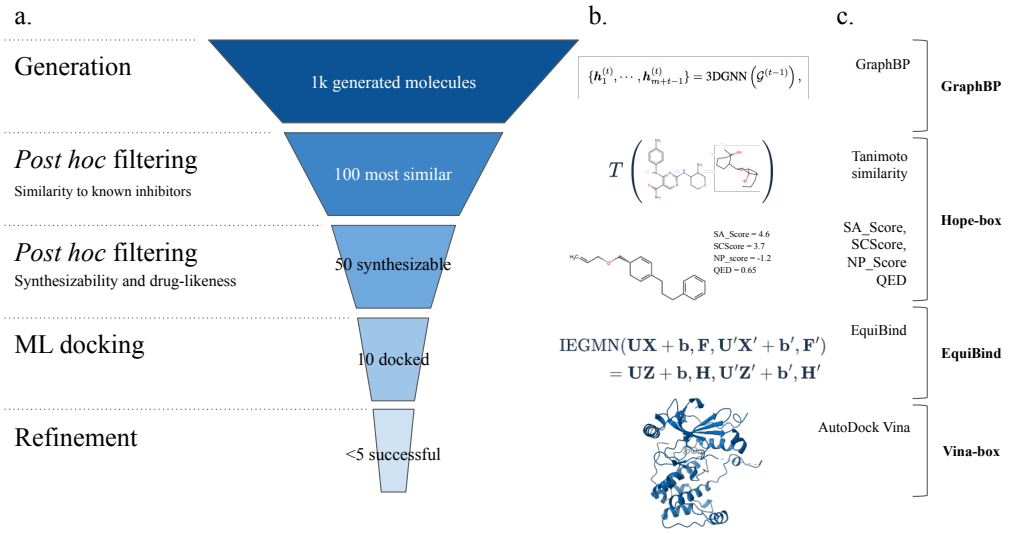


Figure 1: Overview of the computational drug discovery pipeline. (a) Funnel diagram showing the progressive filtering stages: initial set of small-molecule ligands generated by the generative model GraphBP ((Liu et al., 2022)), filtered by similarity to known Aurora kinase inhibitors, synthesizability and drug-likeness, docked to the reference target via the ML-based model Equibind ((Stärk et al., 2022)), and post-processed with AutoDock Vina ((Trott & Olson, 2010)). (b) Representative example showing a generated molecule with its synthesizability and drug-likeness scores, and docked protein-ligand complex (PDB ID: 4AF3). (c) Pipeline submodules (in **bold**) with associated computational methods for each stage.

1 METHODS

1.1 DATA

Benchmarking data. The SMILES strings of all known Aurora kinase inhibitors used in the benchmarking study are obtained from the IUPHAR/BPS Guide to Pharmacology database as a `.csv` file.
Training data. The dataset used for training the generative model is CrossDocked2020 v1.0 (Francoeur et al., 2020), specifically designed for benchmarking machine learning models for predicting protein-ligand binding affinities and poses in structure-based drug design.
Inference. The data used for inference is the three-dimensional structure of the Aurora kinase B-INCENP-VX680 inhibitor complex (PDB ID: 4AF3). Data used for inference has been extracted from the training data so that the two sets are completely non-overlapping.

1.2 MODELS

We choose to implement our Snakemake workflow orchestrator using open-source models and custom evaluation routines for molecular generation, post-hoc evaluation, docking, and binding affinity assessment.
Molecular generative model. GraphBP is a flow-based graph generative model for pocket conditional generation of molecules. GraphBP is fast, flexible, and provides biologically relevant predictions (See Appendix A.2 for details).
Post-hoc filtering. Post-hoc filtering is implemented as similarity calculation and synthesizability assessment via a developed routine that we call `hope-box`. `hope-box` enables easy and fast calculation of the similarity between molecules (See Appendix C.3 for details), calculation of synthetic accessibility (See Appendix D.4.1 for details), and absolute and relative ranking. `hope-box` is used to filter out molecules that do not satisfy evaluation criteria, and prioritize those that represent promising candidates to send to the docking stage.
ML-based docking model. EquiBind is the model of choice for fast, approximate, ML-based docking of the prioritized molecules (See Appendix C.2.2 for details).
Score-based docking. AutoDock Vina is the gold standard score-based model for docking that we choose for final binding affinity calculation of the proposed complexes.

1.3 EVALUATION AND WORKFLOW

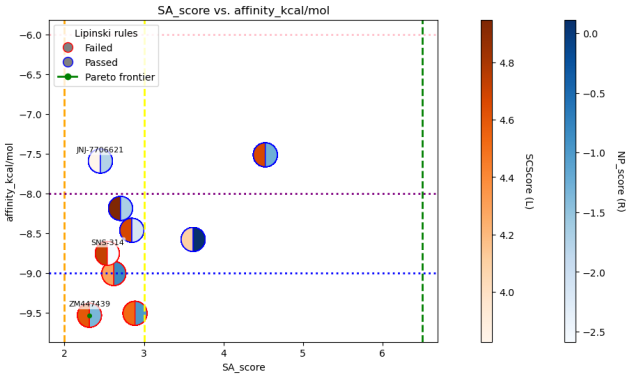
Evaluation. Evaluation is performed post-hoc following criteria developed after completion of the benchmarking study. The evaluation axes are synthetic accessibility, drug-likeness, and binding affinity in kcal/mol. We formulate three tiers that will guide the multi-objective optimization analysis generating the final the Pareto frontier. The values defining the filtering thresholds are shown in Table 1.
Workflow. The workflow is implemented as a Snakemake orchestrator sequentially running molecular generation, post-hoc filtering, ML-based docking, and score-based docking. The framework reflects the steps of the standard drug design pipeline (See Appendix C for details).

2 EXPERIMENTS

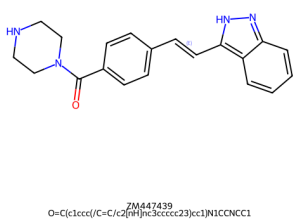
Overview. We present this work as a framework that can handle streamlined de novo drug design of small-molecule drugs conditionally on a chosen biological target. Due to the importance of gaining mechanistic insight in biology, our approach is tested on the case study provided by the Aurora kinase protein family, and targeted to the design of novel ATP-competitive small-molecule inhibitors. As previously mentioned, the ground for our experimental design is a real-life scenario, and the framework strives to widen the accessibility of these tools by providing interpretable predictions in the form of plots, tables, and graphs. Please refer to Appedix D for experimental details and additional results.
Setup. We focus on the case study of Aurora kinase inhibitors. The first experiment is a benchmark study on every known and commercialized Aurora kinase inhibitor; it will serve as a guide and evaluation criterion for the following experiments involving the generation of novel ligands for the Aurora kinases, and will allow us to define some evaluation criteria about the in silico filtering process that will be carried out after the generation step.

Table 1: Performance thresholds defined after benchmarking as filtering tiers.

Filtering tier	Synthesizability scores			Lipinski violations (number)	Vina score (kcal/mol)
	SA_score	SCScore	NP_score		
Permissive	6.5	4.5	0	2	-7
Moderate	4.5	3.5	-1	1	-8
Strict	3	2	-2	0	-9



(a) Pareto frontier of known Aurora kinase B inhibitors. The plot shows the molecules having passed all the stages and reports synthetic accessibility scores, binding affinity, and Lipinski's rules compliance.



(b) The inhibitor ZM447439 is the only molecule in the Pareto frontier. The visualization has been obtained with RDKit.

Figure 2: Benchmarking results.

In all experiments, synthetic accessibility is measured using three proxy scores: SA_Score (Ertl & Schuffenhauer, 2009), SCScore (Coley et al., 2018), and NP_Score (Ertl et al., 2008); structural and functional similarity is calculated using RDKit's (RDKit: open-source cheminformatics and machine learning software) implementation of the Tanimoto similarity[1]; drug-likeness is measured using the quantitative estimate of drug-likeness (QED) score (Bickerton et al., 2012); Lipinski's rule of 5 (Lipinski et al., 2001) compliance is assessed using a script relying on RDKit; docking performance is first established using the confidence score of the ML-based docking model as a proxy (Stärk et al., 2022), and later on estimated using the score-based model AutoDock Vina (Trott & Olson, 2010). Please refer to Appendix D.4.1 for additional details.

Post-hoc evaluation. We choose to use a post-hoc evaluation strategy consisting in the calculation of the chemical and functional similarity, and the evaluation of the synthetic accessibility of the proposed molecules via established proxy scores. We calculate the Pareto frontier of evaluation criteria and provide as output the candidate molecules that compose it.

2.1 Benchmarking against known Aurora kinases inhibitors

We begin our experiments with a benchmark against known and commercialized ATP-competitive Aurora kinase A and B inhibitors retrieved from the IUPHAR/BPS Guide to Pharmacology database[2]. To obtain benchmarking thresholds we run our entire pipeline, end-to-end, with the exclusion of the generative step. Aurora kinase inhibitors are sent through the `hope-box` module, the ML-based docking model, and the score-based docking model. We use the shared setup to comprehensively evaluate the compounds and construct a Pareto frontier populated by molecules with desirable properties. The benchmarking study allows the establishment of performance thresholds for evaluating the molecules proposed by the computational pipeline. Results are reported in Table 1. Figure 2 depicts the results of running the entire pipeline on Aurora kinase B inhibitors.

[1] Also known as Jaccard coefficient.

[2] `https://www.guidetopharmacology.org`

2.2 UNCONSTRAINED, UNBIASED, BINDING POCKET UNCONDITIONAL GENERATION

In this initial experiment, the generative model proposes novel ligand designs for the Aurora kinase target without exploiting any prior information about their desired structural or functional characteristics. The generation process is not biased with any synthetic accessibility score, and the model is not provided with information about the spatial location nor structural biology of the target's binding pocket. Running the pipeline (generation, then post-hoc filtering, then docking) results in a Pareto frontier consisting of molecules that do not exhibit particularly promising scores. Binding affinity values denote especially weak interactions, thus making it improbable for these candidates to be sent further down the drug discovery process. These results show that the chosen molecular generative model can generate chemically plausible molecules, but fails at generating competing molecules when it is not equipped with additional information about the receptor or the chemical space that is desirable to explore; additionally, the results of this initial experiment strengthen the hypothesis about the importance of optimizing for multiple objectives simultaneously to ensure that in silico evaluation proposes promising candidates for following steps of the drug discovery pipeline.

2.3 UNCONSTRAINED, UNBIASED, BINDING POCKET CONDITIONAL GENERATION

In this experiment, the generative model performs inference on the chosen biological target proposing novel ligand designs without any prior information about the desired structural or functional characteristics. However, the model is now equipped with structural and positional information of the target's binding pocket to perform pocket-conditional generation. Running the pipeline results in a Pareto frontier with molecules mostly populating uninteresting areas. Despite displaying some promising scores, the inspection of the three-dimensional structure of the proposed molecules reveals unsuitable conformations.

2.4 CONSTRAINED, UNBIASED, CONDITIONAL GENERATION

In this experiment, we enforce a soft constraint on the generation process by requiring the generative model to propose novel molecular designs such that $n_{\text{min}} \leq i \leq n_{\text{max}}$, where i is the number of atoms of the novel design, n_{min} is the minimum number of atoms in any known Aurora kinase inhibitor, and similarly n_{max} is the maximum number of atoms in any known Aurora kinase inhibitor. No synthetic accessibility score is used to bias the generative model, but the model is given binding pocket information about the target receptor to conditionally generate ligands. After the generation phase, ligands are ranked according to Tanimoto similarity to known Aurora kinase inhibitors, and the top m scoring ligands (with $m = 50$) in synthetic accessibility are docked using the ML-based docking model; subsequently, the top k (with $k = 15$) ligands according to the model's implicit confidence score are sent to a refinement stage performed with the score-based docking method AutoDock Vina. The resulting Pareto frontier is once again composed of < 5 molecules, some of which satisfy the strict filtering criteria, while most of them fall in the moderate-to-permissive tier.

3 DISCUSSION

In conclusion, we have developed and tested a comprehensive and modular computational drug discovery pipeline that successfully addresses the critical fragmentation challenge in the field by integrating diverse state-of-the-art models into a unified workflow. While our systematic evaluation reveals that current GNN-based generative models alone cannot reliably produce viable drug candidates, our pipeline effectively filters and prioritizes generated molecules to identify the most promising candidates for medicinal chemistry evaluation. The modular, plug-and-play architecture implemented through Snakemake democratizes access to sophisticated computational tools while enabling researchers to seamlessly integrate emerging technologies. Although none of the molecules identified in this Aurora kinase case study emerged as particularly promising candidates for immediate clinical advancement, the pipeline's rapid filtering capabilities and flexible architecture establish a robust foundation for future drug discovery efforts. As computational models continue to evolve, this framework provides the necessary infrastructure to harness new developments in generative AI and molecular modeling, ultimately bridging the gap between computational innovation and practical therapeutic development.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M.R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630:493–500, 6 2024. ISSN 14764687. doi: 10.1038/s41586-024-07487-w.

Gustaf Ahdritz, Nazim Bouatta, Christina Floristean, Sachin Kadyan, Qinghui Xia, William Gerecke, Timothy J. O'Donnell, Daniel Berenberg, Ian Fisk, Niccolò Zanichelli, Bo Zhang, Arkadiusz Nowaczynski, Bei Wang, Marta M. Stepniewska-Dziubinska, Shang Zhang, Adegoke Ojewole, Murat Efe Guney, Stella Biderman, Andrew M. Watkins, Stephen Ra, Pablo Ribalta Lorenzo, Lucas Nivon, Brian Weitzner, Yih-En Andrew Ban, Shiyang Chen, Minjia Zhang, Conglong Li, Shuaiwen Leon Song, Yuxiong He, Peter K. Sorger, Emad Mostaque, Zhao Zhang, Richard Bonneau, and Mohammed AlQuraishi. Openfold: retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *Nature Methods*, 21(8): 1514–1524, 2024. doi: 10.1038/s41592-024-02272-z. URL https://doi.org/10.1038/s41592-024-02272-z.

Mohammed Alquraishi et al. AI-driven drug discovery: A comprehensive review. *ACS Omega*, 2025. doi: 10.1021/acsomega.5c00549.

Anthropic. The claude 3 model family: Opus, sonnet, haiku. Technical report, Anthropic, 2024. URL https://www.anthropic.com/news/claude-3-family.

Chris Arter, Luke Trask, Sarah Ward, Sharon Yeoh, and Richard Bayliss. Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors, 8 2022. ISSN 1083351X.

Sajda Ashraf, Kara E. Ranaghan, Christopher J. Woods, Adrian J. Mulholland, and ... Zaheer Ul-Haq. Exploration of the structural requirements of aurora kinase b inhibitors by a combined qsar, modelling and molecular simulation approach. *Scientific Reports*, 11:18707, 2021. doi: 10.1038/s41598-021-97368-3. URL https://www.nature.com/articles/s41598-021-97368-3.

Viraj Bagal, Rishal Aggarwal, P. K. Vinod, and U. Deva Priyakumar. Molgpt: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9):2064–2076, 05 2022. doi: 10.1021/acs.jcim.1c00600. URL https://doi.org/10.1021/acs.jcim.1c00600.

Vassilios Bavetsias and Spiros Linardopoulos. Aurora kinase inhibitors: Current status and outlook. *Frontiers in Oncology*, Volume 5 - 2015, 2015. ISSN 2234-943X. doi: 10.3389/fonc.2015.00278. URL https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2015.00278.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000. doi: 10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

Khushwant S. Bhullar, Naiara Orrego Lagarón, Eileen M. McGowan, Indu Parmar, Amitabh Jha, Basil P. Hubbard, and H. P.Vasantha Rupasinghe. Kinase-targeted cancer therapies: Progress, challenges and future directions. *Molecular Cancer*, 17, 2 2018. ISSN 14764598. doi: 10.1186/s12943-018-0804-2.

Richard G. Bickerton, Gaia V. Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4:90–98, 2012. doi: 10.1038/nchem.1243. URL https://doi.org/10.1038/nchem.1243.

Mar Carmena and William C. Earnshaw. The cellular geography of aurora kinases. *Nature Reviews Molecular Cell Biology*, 4:842–854, 11 2003. ISSN 14710072. doi: 10.1038/nrm1245.

Mar Carmena, William C. Earnshaw, and David M. Glover. The dawn of aurora kinase research: From fly genetics to the clinic. *Frontiers in Cell and Developmental Biology*, Volume 3 - 2015, 2015. ISSN 2296-634X. doi: 10.3389/fcell.2015.00073. URL https://www.frontiersin.org/journals/cell-and-developmental-biology/articles/10.3389/fcell.2015.00073.

Patrizia Carpinelli and Jürgen Moll. Aurora kinase inhibitors: identification and preclinical validation of their biomarkers. *Expert Opinion on Therapeutic Targets*, 12(1):69–80, 2008. doi: 10.1517/14728222.12.1.69. URL https://doi.org/10.1517/14728222.12.1.69. PMID: 18076371.

Jonathan Casper, Matthew L Speir, Brian J Raney, Gerardo Perez, Luis R Nassar, Christopher M Lee, Angie S Hinrichs, Jairo Navarro Gonzalez, Clay Fischer, Mark Diekhans, Hiram Clawson, Anna Benet-Pages, Galt P Barber, Charles J Vaske, Marijke J van Baren, Karen Wang, Yesenia Joanna Puga Rodriguez, Jaidan Ashlyn Jenkins-Kiefer, Megna Chalamala, David Haussler, William James Kent, and Maximilian Haeussler. The ucsc genome browser database: 2026 update. *Nucleic Acids Research*, 54(D1):D1331–D1335, 11 2025. ISSN 1362-4962. doi: 10.1093/nar/gkaf1250. URL https://doi.org/10.1093/nar/gkaf1250.

Shuan Chen and Yousung Jung. Estimating the synthetic accessibility of molecules with building block and reaction-aware sascore. 2024. doi: 10.1186/s13321-024-00879-0. URL https://jcheminf.biomedcentral.com/articles/10.1186/s13321-024-00879-0.

Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Scscore: Synthetic complexity learned from a reaction corpus. *Journal of Chemical Information and Modeling*, 58(2): 252–261, 2018. doi: 10.1021/acs.jcim.7b00622. URL https://doi.org/10.1021/acs.jcim.7b00622. PMID: 29309147.

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.

Miruna Cretu, Charles Harris, Ilia Igashov, Arne Schneuing, Marwin Segler, Bruno Correia, Julien Roy, Emmanuel Bengio, and Pietro Liò. Synflownet: Design of diverse and novel molecules with synthesis constraints, 2025. URL https://arxiv.org/abs/2405.01155.

Altaf A. Dar, Laura W. Goff, Shahana Majid, Jordan Berlin, and Wael El-Rifai. Aurora kinase inhibitors - rising stars in cancer therapeutics? *Molecular Cancer Therapeutics*, 9(2):268–278, 02 2010. ISSN 1535-7163. doi: 10.1158/1535-7163.MCT-09-0765. URL https://doi.org/10.1158/1535-7163.MCT-09-0765.

Christian O. de Groot, Judy E. Hsia, John V. Anzola, Amir Motamedi, Michelle Yoon, Yao Liang Wong, David Jenkins, Hyun J. Lee, Mallory B. Martinez, Robert L. Davis, Timothy C. Gahman, Arshad Desai, and Andrew K. Shiau. A cell biologist's field guide to aurora kinase inhibitors. *Frontiers in Oncology*, Volume 5 - 2015, 2015. ISSN 2234-943X. doi: 10.3389/fonc.2015.00285. URL https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2015.00285.

DeepSeek-AI. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model, 2024. URL https://arxiv.org/abs/2405.04434.

Paolo Di Tommaso, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, and Cedric Notredame. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4):316–319, 2017. doi: 10.1038/nbt.3820. URL https://doi.org/10.1038/nbt.3820.

Hannah Dowden and Jamie Munro. Trends in clinical success rates and therapeutic focus. *Nature Reviews Drug Discovery*, 18:495–496, 2019. doi: 10.1038/d41573-019-00074-z.

Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, 1(1):8, 2009. doi: 10.1186/1758-2946-1-8. URL https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-1-8.

Peter Ertl, Silvio Roggo, and Ansgar Schuffenhauer. Natural product-likeness score and its application for prioritization of compound libraries. *Journal of Chemical Information and Modeling*, 48(1):68–74, 2008. doi: 10.1021/ci700286x. URL https://doi.org/10.1021/ci700286x. PMID: 18034468.

Paul G. Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B. Iovanisci, Ian Snyder, and David R. Koes. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling*, 60(9): 4200–4215, 2020. doi: 10.1021/acs.jcim.0c00411. URL https://pubs.acs.org/doi/10.1021/acs.jcim.0c00411.

Octavian-Eugen Ganea, Xinyuan Huang, Charlotte Bunne, Yatao Bian, Regina Barzilay, Tommi Jaakkola, and Andreas Krause. Independent se(3)-equivariant models for end-to-end rigid protein docking, 2022. URL https://arxiv.org/abs/2111.07786.

Amit Gangwal and Antonio Lavecchia. Artificial intelligence in preclinical research: enhancing digital twins and organ-on-chip to reduce animal testing. *Drug Discovery Today*, 30(5):104360, 2025. ISSN 1359-6446. doi: https://doi.org/10.1016/j.drudis.2025.104360. URL https://www.sciencedirect.com/science/article/pii/S135964462500073X.

Wenhao Gao and Connor W. Coley. The synthesizability of molecules proposed by generative models. *Journal of Chemical Information and Modeling*, 60(12):5714–5723, 2020. doi: 10.1021/acs.jcim.0c00174. URL https://doi.org/10.1021/acs.jcim.0c00174. PMID: 32250616.

David M Glover, Mark H Leibowitz, Doris A McLean, and Huw Parry. Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell*, 81(1):95–105, 1995. ISSN 0092-8674. doi: https://doi.org/10.1016/0092-8674(95)90374-7. URL https://www.sciencedirect.com/science/article/pii/0092867495903747.

David S. Goodsell and Arthur J. Olson. Automated docking of substrates to proteins by simulated annealing. *Proteins: Structure, Function, and Bioinformatics*, 8(3):195–202, 1990. doi: https://doi.org/10.1002/prot.340080302. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.340080302.

Simon D Harding, Jane F Armstrong, Elena Faccenda, Christopher Southan, Stephen P H Alexander, Anthony P Davenport, Michael Spedding, and Jamie A Davies. The iuphar/bps guide to pharmacology in 2026. *Nucleic Acids Research*, 54(D1):D1446–D1456, 10 2025. ISSN 1362-4962. doi: 10.1093/nar/gkaf1067. URL https://doi.org/10.1093/nar/gkaf1067.

Emiel Hoogeboom, Victor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3d, 2022. URL https://arxiv.org/abs/2203.17003.

Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Machine Learning: Science and Technology*, 3 (1):015022, jan 2022. doi: 10.1088/2632-2153/ac3ffb. URL https://doi.org/10.1088/2632-2153/ac3ffb.

Chang Jin, Siyuan Guo, Shuigeng Zhou, and Jihong Guan. Effective and explainable molecular property prediction by chain-of-thought enabled large language models and multi-modal molecular information fusion. *Journal of Chemical Information and Modeling*, 65(11):5438–5455, 2025. doi: 10.1021/acs.jcim.5c00577. URL https://doi.org/10.1021/acs.jcim.5c00577. PMID: 40392109.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation, 2019. URL https://arxiv.org/abs/1802.04364.

Xue-Li Jing and Shi-Wu Chen. Aurora kinase inhibitors: a patent review (2014-2020). *Expert Opinion on Therapeutic Patents*, 31(7):625–643, 2021. doi: 10.1080/13543776.2021.1890027. URL https://doi.org/10.1080/13543776.2021.1890027. PMID: 33573401.

Hiroshi Katayama, William R. Brinkley, and Subrata Sen. The aurora kinases: Role in cell transformation and tumorigenesis. *Cancer and Metastasis Reviews*, 22(4):451–464, 2003. doi: 10.1023/A:1023789416385. URL https://doi.org/10.1023/A:1023789416385.

Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024, October 2020. ISSN 2632-2153. doi: 10.1088/2632-2153/aba947. URL http://dx.doi.org/10.1088/2632-2153/aba947.

Rohith Krishna, Jue Wang, Woody Ahern, Pascal Sturmfels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, Felix S. Morey-Burrows, Ivan Anishchenko, Ian R. Humphreys, Ryan McHugh, Dionne Vafeados, Xinting Li, George A. Sutherland, Andrew Hitchcock, C. Neil Hunter, Alex Kang, Evans Brackenbrough, Asim K. Bera, Minkyung Baek, Frank DiMaio, and David Baker. Generalized biomolecular modeling and design with rosettafold all-atom. *Science*, 384(6693): eadl2528, 2024. doi: 10.1126/science.adl2528. URL https://www.science.org/doi/abs/10.1126/science.adl2528.

Irwin D. Kuntz, Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, and Thomas E. Ferrin. A geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, 161(2):269–288, 1982. ISSN 0022-2836. doi: https://doi.org/10.1016/0022-2836(82)90153-X. URL https://www.sciencedirect.com/science/article/pii/002228368290153X.

Nicholas Kwiatkowski, Xianming Deng, Jinhua Wang, Li Tan, Fabrizio Villa, Stefano Santaguida, Hsiao-Chun Huang, Tim Mitchison, Andrea Musacchio, and Nathanael Gray. Selective aurora kinase inhibitors identified using a taxol-induced checkpoint sensitivity screen. *ACS Chemical Biology*, 7(1):185–196, 2012. doi: 10.1021/cb200305u. URL https://doi.org/10.1021/cb200305u. PMID: 21992004.

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning, 2019. URL https://arxiv.org/abs/1910.09700.

Seung Yun Lee, Da Young Lee, Ji Hyeop Kang, Jae Won Jeong, Jae Hyeon Kim, Hyun Woo Kim, Dong Hoon Oh, Jun-Mo Kim, Shin-Jae Rhim, Gap-Don Kim, Hyeong Sang Kim, Young Dal Jang, Yeonhwa Park, and Sun Jin Hur. Alternative experimental approaches to reduce animal use in biomedical studies. *Journal of Drug Delivery Science and Technology*, 68:103131, 2022. ISSN 1773-2247. doi: https://doi.org/10.1016/j.jddst.2022.103131. URL https://www.sciencedirect.com/science/article/pii/S1773224722000405.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574. URL https://www.science.org/doi/abs/10.1126/science.ade2574.

Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1pii of original article: S0169-409x(96)00423-1. the article was originally published in advanced drug delivery reviews 23 (1997) 3–25.1. *Advanced Drug Delivery Reviews*, 46(1):3–26, 2001. ISSN 0169-409X. doi: https://doi.org/10.1016/S0169-409X(00)00129-0. URL https://www.sciencedirect.com/science/article/pii/S0169409X00001290. Special issue dedicated to Dr. Eric Tomlinson, Advanced Drug Delivery Reviews, A Selection of the Most Highly Cited Articles, 1991-1998.

Cheng-Hao Liu, Maksym Korablyov, Stanislaw Jastrzebski, Pawel Wlodarczyk-Pruszynski, Yoshua Bengio, and Marwin H. S. Segler. Retrognn: Approximating retrosynthesis by graph neural networks for de novo drug design, 2020. URL https://arxiv.org/abs/2011.13042.

Gang Liu, Jiaxin Xu, Tengfei Luo, and Meng Jiang. Graph diffusion transformers for multi-conditional molecular generation, 2024. URL https://arxiv.org/abs/2401.13858.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. Generating 3d molecules for target protein binding, 2022. URL https://arxiv.org/abs/2204.09410.

Hannes H. Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H. Mervin, and Ola Engkvist. Reinvent 4: Modern ai–driven generative molecule design. *Journal of Cheminformatics*, 16, 12 2024. ISSN 17582946. doi: 10.1186/s13321-024-00812-5.

Tomotoshi Marumoto, Dongwei Zhang, and Hideyuki Saya. Aurora a-a guardian of poles. *Nature Reviews Cancer*, 5(1):42–50, 2005. doi: 10.1038/nrc1526. URL https://doi.org/10.1038/nrc1526.

Elaine C. Meng, Thomas D. Goddard, Eric F. Pettersen, Greg S. Couch, Zach J. Pearson, John H. Morris, and Thomas E. Ferrin. Ucsf chimerax: Tools for structure building and analysis. *Protein Science*, 32(11):e4792, 2023. doi: https://doi.org/10.1002/pro.4792. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4792.

Patrick Meraldi, Reiko Honda, and Erich A. Nigg. Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Current Opinion in Genetics and Development*, 14:29–36, 2004. ISSN 0959437X. doi: 10.1016/j.gde.2003.11.006.

Garrett M. Morris, Ruth Huey, William Lindstrom, Michel F. Sanner, Richard K. Belew, David S. Goodsell, and Arthur J. Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16):2785–2791, 2009. doi: https://doi.org/10.1002/jcc.21256. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21256.

Felix Mölder, Kim Philipp Jablonski, Brice Letcher, Michael B. Hall, Christopher H. Tomkins-Tinch, Vanessa Sochat, Jan Forster, Soohyun Lee, Sven O. Twardziok, Alexander Kanitz, Andreas Wilm, Manuel Holtgrewe, Sven Rahmann, Sven Nahnsen, and Johannes Köster. Sustainable data analysis with snakemake. *F1000Research*, 10:33, 1 2021. doi: 10.12688/f1000research.29032.1.

OpenAI. Gpt-4o system card, 2024. URL https://arxiv.org/abs/2410.21276.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707. URL https://www.biorxiv.org/content/early/2025/06/18/2025.06.14.659707.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2mol: Efficient molecular sampling based on 3d protein pockets, 2025. URL https://arxiv.org/abs/2205.07249.

Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004. doi: https://doi.org/10.1002/jcc.20084. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20084.

Jianbo Qiao, Junru Jin, Ding Wang, Saisai Teng, Junyu Zhang, Xuetong Yang, Yuhang Liu, Yu Wang, Lizhen Cui, Quan Zou, Ran Su, and Leyi Wei. A self-conformation-aware pre-training framework for molecular property prediction with substructure interpretability. *Nature Communications*, 16, 2025. doi: 10.1038/s41467-025-59634-0. URL https://www.nature.com/articles/s41467-025-59634-0.

RDKit: open-source cheminformatics and machine learning software. Rdkit.

Arman A. Sadybekov and Vsevolod Katritch. Computational approaches streamlining drug discovery. *Nature*, 616:673–685, 2023. doi: 10.1038/s41586-023-05905-z. URL https://www.nature.com/articles/s41586-023-05905-z.

Tatsuya Sagawa and Ryosuke Kojima. Reactiont5: a large-scale pre-trained model towards application of limited reaction data, 2023. URL https://arxiv.org/abs/2311.06708.

Lakshidaa Saigiridharan, Alan Kai Hassen, Helen Lai, Paula Torren-Peraire, Ola Engkvist, Samuel Genheden, et al. Aizynthfinder 4.0: developments based on learnings from 3 years of industrial application. *Journal of Cheminformatics*, 16(1):57, 2024. doi: 10.1186/s13321-024-00860-x. URL https://jcheminf.biomedcentral.com/articles/10.1186/s13321-024-00860-x.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L. Blundell, Pietro Lio, Max Welling, Michael Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12):899–909, 2024. doi: 10.1038/s43588-024-00737-x. URL https://doi.org/10.1038/s43588-024-00737-x.

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

Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A. Hunter, Costas Bekas, and Alpha A. Lee. Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(9):1572–1583, 2019. doi: 10.1021/acscentsci.9b00576. URL https://doi.org/10.1021/acscentsci.9b00576. PMID: 31572784.

Aylin Sertkaya, Trinidad Beleche, Amber Jessup, and Benjamin D. Sommers. Costs of drug development and research and development intensity in the us, 2000-2018. *JAMA Network Open*, 7(6):e2415445–e2415445, 06 2024. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2024.15445. URL https://doi.org/10.1001/jamanetworkopen.2024.15445.

Tony Shen, Seonghwan Seo, Ross Irwin, Kieran Didi, Simon Olsson, Woo Youn Kim, and Martin Ester. Compositional flows for 3d molecule and synthesis pathway co-design, 2025. URL https://arxiv.org/abs/2504.08051.

Martin Simonovsky and Nikos Komodakis. Graphvae: Towards generation of small graphs using variational autoencoders, 2018. URL https://arxiv.org/abs/1802.03480.

Hannes Stärk, Octavian-Eugen Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. Equibind: Geometric deep learning for drug binding structure prediction. 2 2022. URL http://arxiv.org/abs/2202.05146.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *bioRxiv*, 2024. doi: 10.1101/2024.10.10.615955. URL https://www.biorxiv.org/content/early/2024/10/11/2024.10.10.615955.

Oleg Trott and Arthur J. Olson. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: https://doi.org/10.1002/jcc.21334. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21334.

Zhengkai Tu, Sourabh J. Choure, Mun Hong Fong, Jihye Roh, Itai Levin, Kevin Yu, Joonyoung F. Joung, Nathan Morgan, Shih-Cheng Li, Xiaoqi Sun, Huiqian Lin, Mark Murnin, Jordan P. Liles, Thomas J. Struble, Michael E. Fortunato, Mengjie Liu, William H. Green, Klavs F. Jensen, and Connor W. Coley. Askcos: Open-source, data-driven synthesis planning. *Accounts of Chemical Research*, 58(11):1764–1775, 2025. doi: 10.1021/acs.accounts.5c00155. URL https://doi.org/10.1021/acs.accounts.5c00155. PMID: 40397546.

U.S. Food and Drug Administration. FDA announces plan to phase out animal testing requirement for monoclonal antibodies and other drugs, January 2025a. URL https://www.fda.gov/news-events/press-announcements/fda-announces-plan-phase-out-animal-testing-requirement-monoclonal-antibodies-and-

U.S. Food and Drug Administration. Roadmap to reducing animal testing in preclinical safety studies, 2025b. URL https://www.fda.gov/files/newsroom/published/roadmap_to_reducing_animal_testing_in_preclinical_safety_studies.pdf.

Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, et al. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6):463–477, 2019. doi: 10.1038/s41573-019-0024-5.

Prerna Vats, Chainsee Saini, Bhavika Baweja, Sandeep K. Srivastava, Ashok Kumar, Atar Singh Kushwah, and Rajeev Nema. Aurora kinases signaling in cancer: from molecular perception to targeted therapies. *Molecular Cancer*, 24:180, 2025. doi: 10.1186/s12943-025-02353-3. URL https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-025-02353-3.

Divya Vemula, Perka Jayasurya, Varthiya Sushmitha, Yethirajula Naveen Kumar, and Vasundhra Bhandari. Cadd, ai and ml in drug discovery: A comprehensive review. *European Journal of Pharmaceutical Sciences*, 181:106324, 2023. ISSN 0928-0987. doi: https://doi.org/10.1016/j.ejps.2022.106324. URL https://www.sciencedirect.com/science/article/pii/S0928098722002093.

Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. Digress: Discrete denoising diffusion for graph generation, 2023. URL https://arxiv.org/abs/2209.14734.

Min Yan, Chunli Wang, Bin He, Mengying Yang, Mengying Tong, Zijie Long, Bing Liu, Fei Peng, Lingzhi Xu, Yan Zhang, Dapeng Liang, Haixin Lei, Sen Subrata, Keith W. Kelley, Eric W.-F. Lam, Bilian Jin, and Quentin Liu. Aurora-a kinase: A potent oncogene and target for cancer therapy. *Medicinal Research Reviews*, 36(6):1036–1079, 2016. doi: https://doi.org/10.1002/med.21399. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/med.21399.

ADDITIONAL INFORMATION

DATA AVAILABILITY

All data used in this study is publicly available. Please refer to Appendix B.1 for the links to access the data.

CODE AVAILABILITY

The source code to reproduce every experiment in this work is publicly available at the following anonymized repository: `https://anonymous.4open.science/r/drug-design-pipeline-BF4F`.

CO2 EMISSION RELATED TO EXPERIMENTS

The experiments were carried out using GPUs made available by *anonymized funding institution* which has an approximate carbon efficiency of 0.432 kgCO^2eq/kWh. A cumulative of 1380 hours of computation was performed on hardware of type A100 PCIe 40/80GB (TDP of 250W). The total emissions are estimated to be 149.04 kgCO^2eq of which 0 percent were directly offset. The aforementioned carbon emission roughly corresponds to 602 Km driven by an average ICE car[3], 74.6 Kgs of coal burned[4], and 2.48 tree seedlings sequestering carbon for 10 years[5]. Estimates were performed using the Machine Learning Impact calculator presented in Lacoste et al. (2019).

AUTHOR CONTRIBUTIONS

Anonymous Author 1 conceptualized the study, designed and conducted the experiments, analyzed the data, and wrote the manuscript.
Anonymous Author 2 provided feedback on the manuscript.
All authors approved the final version.

ACKNOWLEDGMENTS

The authors acknowledge the use of conversational agents ChatGPT (OpenAI, 2024), DeepSeek (DeepSeek-AI, 2024), and Claude Sonnet (Anthropic, 2024) to assist in summarization of content, language editing and refinement.

EXPERIMENTAL SETTING

Additional information about the experimental setting can be found in the following sections of the Appendix.

[3] Greenhouse Gas Equivalencies Calculator
[4] Ibid., LBs coal
[5] Ibid., Seedlings

A RELATED WORKS

A.1 CADD AND DE NOVO DRUG DISCOVERY

Computer-aided drug discovery (CADD) leverages computational tools such as molecular modeling, machine learning, and bioinformatics to streamline drug development. By simulating drug-target interactions (e.g., docking, virtual screening), CADD identifies potential drug candidates, optimizes their properties, and reduces experimental trial-and-error, saving time and costs. De novo drug design is a crucial and often early-phase strategy within the broader landscape of CADD. It is primarily a hit identification strategy that refers to designing novel molecules from scratch (de novo) that could become potential drugs, without relying on known active compounds. The standard pipeline for de novo drug discovery is defined by the following steps: (1) Target identification and validation; (2) Structural biology and target characterization; (3) *De novo* ligand design; (3.1) Binding site analysis; (3.2) Choice of granularity; (3.3) Design strategy; (3.4) Virtual screening and scoring; (4) In silico filtering; (5) Synthesis and validation; (6) Hit-to-lead optimization (Vemula et al., 2023; Sadybekov & Katritch, 2023).

A.2 GRAPH GENERATIVE MODELS

Graph generative models have emerged as powerful tools for molecular generation in drug discovery, leveraging the natural graph structure of molecules to generate novel compounds with desired properties. These models usually represent molecules as graphs where atoms are nodes and bonds are edges (See Appendix B.2 for details), capturing complex molecular topology and chemistry. Graph Neural Networks (GNNs) like GraphVAE (Simonovsky & Komodakis, 2018) and Junction Tree Variational Autoencoders (JT-VAE) (Jin et al., 2019) learn continuous latent representations of molecular graphs, allowing for smooth interpolation in chemical space and the generation of valid molecules through decoding from the latent space; these approaches excel at preserving local molecular motifs and generating chemically plausible structures. Diffusion models, such as DiGress (Vignac et al., 2023) and EDM (Equivariant Diffusion Models) (Hoogeboom et al., 2022), have shown remarkable success by learning to reverse a gradual noising process applied to molecular graphs; these models can generate high-quality molecules by iteratively denoising random graphs, often producing more diverse and novel structures compared to VAE-based approaches. The probabilistic nature of diffusion models allows for better control over the generation process and improved sample quality. Transformer-based models like SELFIES-based generators (Krenn et al., 2020) and Graph-Transformer (Liu et al., 2024; Loeffler et al., 2024), adapt the attention mechanism to molecular graphs, enabling the model to capture long-range dependencies within molecules; these models can process molecular representations as sequences (SMILES/SELFIES (See Appendix B.2 for details)) or directly operate on graph structures, benefiting from the transformer's ability to model complex relationships and generate coherent molecular structures. Each approach offers distinct advantages: GNNs provide interpretable latent spaces, diffusion models excel in sample quality and diversity, while transformers leverage powerful attention mechanisms for capturing molecular complexity. The choice of model often depends on specific requirements such as generation speed, molecular validity, and the ability to incorporate domain-specific constraints for lead optimization.

A.3 DOCKING MODELS

Molecular docking is a computational method in structural biology that predicts the preferred orientation and binding affinity of one molecule when it binds to another to form a stable complex. Traditional docking approaches like AutoDock (Goodsell & Olson, 1990; Morris et al., 2009) and DOCK (Kuntz et al., 1982) use physics-based scoring functions combined with search algorithms to explore the conformational space and identify energetically favorable binding poses. However, these methods can be especially computationally expensive. Recent advances in machine learning have revolutionized the docking field. Graph neural networks (GNNs) have emerged as particularly powerful tools because they naturally represent molecular structures as graphs, where atoms are nodes and bonds are edges. EquiBind (see Appendix C.2.2 for details) uses GNN architectures to perform fast, direct prediction of protein-ligand complexes without requiring expensive sampling, achieving significant speedups over traditional methods. For protein-protein docking, models like EquiDock (Ganea et al., 2022) extend these principles to predict the transformation that aligns two

protein structures into their bound complex. DiffDock (Corso et al., 2023) introduced diffusion models to the docking problem, treating it as a generative task where the model learns to denoise random ligand poses into correct binding configurations through a diffusion process. Building on the success of protein language models like ESM-2 (Lin et al., 2023), which learn representations from protein sequences, researchers have developed models that incorporate structural information. RoseTTAFold All-Atom (Krishna et al., 2024) and Chai-1 (team et al., 2024) use principles similar to AlphaFold (Abramson et al., 2024) but extend them to predict complexes involving proteins, small molecules, and nucleic acids. These models leverage the transformer architecture to capture long-range dependencies and evolutionary information encoded in sequences. The integration of sequence-based language models with geometric deep learning represents a promising direction, as it combines the rich evolutionary information captured by language models with the physical constraints of 3D structure, potentially enabling more accurate prediction of binding poses and dynamics in biological complexes.

A.4 CHEMICAL SYNTHESIZABILITY

In computer-aided drug design (CADD), post hoc filtering of generated molecules is a common strategy to assess their synthetic accessibility using proxy scores and computer-assisted synthesis planning (CASP) tools. Post hoc filtering involves applying a CASP tool to eliminate unsynthesizable molecules from an unbiased generation process. Evaluating synthetic accessibility is crucial in overcoming a major bottleneck in CADD: even promising candidate molecules are of limited value if they cannot be feasibly synthesized in a laboratory. While experienced medicinal chemists can often estimate synthesizability through visual inspection, computational proxy scores provide a rapid and scalable alternative, enabling efficient prioritization of viable candidates. Despite its importance, chemical synthesizability is often sidelined in early-stage drug discovery, leading to downstream attrition in the development pipeline. To address this, several computational metrics have been proposed: SAscore (Synthetic Accessibility Score) (Ertl & Schuffenhauer, 2009) estimates synthetic complexity based on molecular fragments and complexity penalties; SCscore (Synthetic Complexity Score) (Coley et al., 2018) is a machine learning-based metric trained on retrosynthetic reactions to predict synthetic difficulty; NPscore (Natural Product-likeness Score) (Ertl et al., 2008) quantifies the likelihood of a molecule being natural product-like, indirectly reflecting synthetic feasibility. It is important to note that these scores serve as proxies and do not always perfectly correlate with real-world synthetic accessibility, but integrating such filters into CADD pipelines can significantly enhance efficiency by prioritizing more viable compounds early in the process. Beyond simple scoring functions, advanced tools such as ASKCOS (Tu et al., 2025) and AiZynthFinder (Saigiridharan et al., 2024) leverage retrosynthetic analysis and reaction prediction to provide more refined assessments of synthetic pathways, further bridging the gap between in silico design and laboratory synthesis. Rather than relying solely on post hoc filtering, recent approaches have sought to integrate synthesizability directly into the molecular generation process itself. Graph-based generative models like REINVENT (Loeffler et al., 2024) and MolGPT (Bagal et al., 2022) can be guided by reward functions that incorporate synthetic accessibility scores, steering generation toward more synthetically feasible regions of chemical space. Diffusion models for molecular design, such as DiffSBDD (Schneuing et al., 2024) and Pocket2Mol (Peng et al., 2025), can similarly be conditioned on synthesizability constraints during the denoising process. More sophisticated approaches employ reaction-based generation, where molecules are built through sequences of valid chemical reactions, inherently ensuring synthetic viability. Tools like Molecular Transformer (Schwaller et al., 2019) and RetroGNN (Liu et al., 2020) use transformer and graph neural network architectures to predict reaction outcomes and retrosynthetic routes, enabling more chemically grounded generation. The integration of large language models trained on chemical reactions, such as ChemFormer (Irwin et al., 2022) and ReactionT5 (Sagawa & Kojima, 2023), has further advanced this field by treating synthesis planning as a sequence-to-sequence translation problem, achieving impressive accuracy in predicting both forward reactions and retrosynthetic disconnections.

A.5 WORKFLOW ORCHESTRATORS

Workflow orchestrators in drug discovery are computational frameworks designed to automate, integrate, and streamline the complex, multi-step processes of identifying and optimizing therapeutic candidates. Creating a workflow orchestrator involves designing a system that can coordinate diverse computational tools, ranging from molecular generation and docking to property prediction and syn-

thesis planning, into cohesive, automated pipelines. These orchestrators must handle heterogeneous data formats, manage dependencies between sequential and parallel tasks, enable reproducibility, and often incorporate feedback loops where results from one stage inform earlier steps. The goal is to reduce manual intervention, accelerate discovery timelines, and ensure that computational experiments are systematic and scalable. Modern orchestrators increasingly incorporate machine learning components that can learn from previous iterations to optimize the search through chemical space more intelligently. The tasks typically addressed by drug discovery workflow orchestrators include hit identification, lead optimization, virtual screening campaigns, and de novo molecular design. A common workflow might start with target structure preparation, followed by large-scale virtual screening using docking or machine learning-based binding affinity prediction, filtering by ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties, assessment of synthetic accessibility, and finally prioritization of candidates for experimental validation (See Appendix F for the details of our own implementation of the orchestrator). More sophisticated orchestrators implement active learning loops where experimental feedback is used to retrain predictive models, or multi-objective optimization where molecules are simultaneously optimized for potency, selectivity, druglikeness, and synthesizability. Some workflows integrate retrosynthetic planning tools to ensure that promising candidates have viable synthetic routes before they are suggested for synthesis.

Several notable platforms and publications have emerged in recent years addressing workflow orchestration in drug discovery. AstraZeneca's REINVENT (Loeffler et al., 2024) represents one of the earlier examples of an integrated framework for de novo drug design that combines molecular generation with multi-parameter optimization using reinforcement learning. AlphaFold's (Abramson et al., 2024) integration into drug discovery pipelines has prompted the development of orchestrators that combine structure prediction with virtual screening, as seen in platforms like OpenFold (Ahdritz et al., 2024) workflows and various commercial implementations. Plenty of proprietary platforms have also developed their own orchestrators. The emerging trend involves creating modular, interoperable systems where researchers can plug in state-of-the-art models for each component while maintaining the overall orchestration logic, often leveraging workflow management systems like Nextflow (Di Tommaso et al., 2017), Snakemake (Mölder et al., 2021), or cloud-based platforms to handle the computational infrastructure.

B DATA

B.1 DATASETS

B.1.1 IUPHAR/BPS GUIDE TO PHARMACOLOGY

The IUPHAR/BPS Guide to Pharmacology (GtoPdb) (Harding et al., 2025) is a comprehensive, expert-curated database that serves as an authoritative resource for pharmacological information on drug targets and their ligands. Developed through a collaboration between the International Union of Basic and Clinical Pharmacology (IUPHAR) and the British Pharmacological Society (BPS), GtoPdb provides detailed, quantitative information on the relationships between therapeutic targets and their endogenous and synthetic ligands. The database contains extensively annotated data on binding affinities, functional activities, and pharmacological parameters such as Ki, IC50, EC50, and Kd values, all supported by literature references and expert reviewed. Each entry includes structural information, nomenclature, tissue distribution, physiological function, and clinical relevance of targets, along with comprehensive ligand data covering approved drugs, experimental compounds, endogenous substances, and tool compounds. GtoPdb also provides information on disease associations, making it valuable for target identification and validation in drug discovery. The database emphasizes quantitative pharmacology and provides standardized, quality-controlled data that enables researchers to make informed decisions about target-ligand interactions. GtoPdb is freely accessible online at this link.
We obtain the data for our study navigating the database as follows: `Home > Targets > Enzymes > Kinases (EC 2.7.x.x) > Other protein kinases > Aurora kinase (Aur) family > aurora kinase B`. We then download the `.csv` file containing cumulative information about the target's inhibitors and measured characteristics.

B.1.2 CROSSDOCKED2020

The Crossdocked2020 dataset (Francoeur et al., 2020) was designed to benchmark machine learning models for predicting protein-ligand binding affinities and poses in structure-based drug design. The dataset contains approximately 22.5 million docked poses of ligands across various protein binding pockets. Each sample includes a protein-ligand complex with associated docking pose and binding affinity data (Ki, Kd, IC50). Ligands were docked into multiple similar binding pockets from the Protein Data Bank (PDB) to generate cross-docked poses. It provides standardized splits for clustered cross-validation, ensuring that similar proteins are grouped together to prevent data leakage. The dataset is suitable for training and evaluating models in tasks like binding affinity prediction, pose selection, and virtual screening, and includes a mixture of both high-quality and low-quality docking poses to enhance model robustness. CrossDocked2020 aims to provide a standardized benchmark for comparing different machine learning models in drug discovery. We use the CrossDocked 2020 dataset v1.0 available here[6] to retrain our molecular generative model.

B.1.3 PROTEIN DATA BANK (PDB)

The Protein Data Bank (PDB) (Berman et al., 2000) is the most well-established global repository for three-dimensional structural data of biological macromolecules, including proteins, nucleic acids, and complex assemblies. Since its creation, the PDB has grown to contain over 200,000 experimentally determined structures, making it an indispensable resource for structural biology, drug discovery, and molecular modeling. Structures deposited in the PDB are primarily determined through X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM). Each PDB entry contains atomic coordinates, experimental metadata, ligand information, binding site details, and quality metrics that allow researchers to assess the reliability of structural models. The database provides critical insights into protein function, molecular mechanisms, and drug-target interactions at atomic resolution. For drug discovery specifically, the PDB enables structure-based design approaches by providing templates for homology modeling, targets for virtual screening campaigns, and structural information about protein-ligand complexes that reveal binding modes and facilitate lead optimization. The PDB is freely accessible online to the global scientific community at this link, where it has become the foundational infrastructure for

[6] `https://bits.csb.pitt.edu/files/crossdock2020/v1.0/`

computational biology, enabling advances in fields ranging from enzyme engineering to vaccine design and serving as training data for modern machine learning approaches.
From the PDB repository we extract the `.pdb` file containing the 3D structure of the Aurora kinase B in complex with INCENP and inhibitor VX-680 (tozasertib) (PDB ID: 4AF3), and the FASTA file containing the protein sequence of the complex.

B.2 DATA REPRESENTATION

B.2.1 MOLECULAR REPRESENTATION

SMILES. The Simplified Molecular Input Line Entry System (SMILES) is a line notation for describing the structure of chemical species using short ASCII strings. SMILES strings can be imported by most molecule editors for conversion back into two-dimensional drawings or three-dimensional models of the molecules.

SELFIES. The Self-Referencing Embedded Strings (SELFIES) a 100% robust, string-based molecular representation designed for machine learning, ensuring every generated string translates to a valid molecule. Unlike SMILES, SELFIES use a unique grammar to avoid syntax errors.

Graphs. Graphs are a straightforward representation of molecular structures. Following the data representation strategies in Liu et al. (2022) and Stärk et al. (2022), we represent the 3D geometry of a molecule (i.e., a ligand) as $\mathcal{M} = \{(\mathbf{a}_i, \mathbf{r}_i)\}_{i=1}^{n}$ and the corresponding binding site of a protein (i.e., a receptor) as $\mathcal{P} = \{(\mathbf{b}_j, \mathbf{s}_j)\}_{j=1}^{m}$. n and m denote the numbers of atoms in the molecule and in the binding site, respectively. $\mathbf{a}_i \in \{0,1\}^p$ is the one-hot vector indicating the atom type of the i-th atom in the molecule, and $\mathbf{r}_i \in \mathbb{R}^3$ is its 3D Cartesian coordinate. Similarly, the atom type and the coordinate of the j-th atom in the binding site are denoted as one-hot vector $\mathbf{b}_j \in \{0,1\}^q$ and $\mathbf{s}_j \in \mathbb{R}^3$. p and q represent the total numbers of atom types in molecules and in binding sites, respectively, and they can be obtained from the statistics of the training set. GraphBP considers the problem of generating 3D molecules in the given binding site, thus is trained to learn the conditional distribution $p(\mathcal{M}|\mathcal{P})$ of observed protein-ligand pairs.

B.2.2 PROTEIN REPRESENTATION

Protein sequence. The protein sequence, or, more technically, the protein's primary structure, is the linear sequence of amino acids in a peptide or protein. By convention, the primary structure of a protein is reported starting from the amino-terminal (N) end to the carboxyl-terminal (C) end.

Graphs. Graphs are an amenable representation of protein secondary and tertiary structure. Following Stärk et al. (2022), we define protein graphs as objects like $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where for proteins $\mathcal{V}$ is α-carbons and $\mathcal{E}$ is the k-nearest neighbour (k-NN) of nodes.

B.3 MOLECULAR VISUALIZATION SOFTWARES

Molecular visualization softwares are essential for understanding three-dimensional structures of biomolecules, enabling the exploration of protein architectures, binding sites, and molecular interactions that are critical for drug discovery and structural biology. Among the most widely used programs are PyMOL (Schrödinger, LLC, 2015), known for publication-quality rendering, and the UCSF Chimera suite (Pettersen et al., 2004). Chimera's successor, ChimeraX (Meng et al., 2023), provides significantly improved performance for visualizing massive assemblies like viruses and ribosomes, supports virtual reality interfaces, and incorporates advanced features for density map fitting, molecular dynamics trajectory visualization, and integration with AlphaFold predictions. Both Chimera and ChimeraX are freely available, extensively documented, and scriptable via Python, making them accessible to both novice users and computational experts. Their intuitive interfaces combined with powerful analytical capabilities have made them indispensable tools in structural pharmacology and computational drug design.
In this work, molecular graphics and analyses are performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

C COMPUTER-AIDED DRUG DISCOVERY PIPELINE

C.1 TARGET IDENTIFICATION AND VALIDATION

The goal of this step is to find a biological target (e.g. a protein, enzyme, receptor) associated with a disease. Aurora kinases are compelling yet complex targets that exemplify both the promise and challenges of modern drug discovery. Aurora kinases play essential roles in cancer progression, with different members of the kinase family being overexpressed in different cancer types (Glover et al., 1995; Marumoto et al., 2005; Dar et al., 2010; Katayama et al., 2003; Yan et al., 2016). Despite decades of research (Carmena et al., 2015), existing Aurora kinase inhibitors face significant limitations, including lack of isoform selectivity leading to off-target toxicity, emergence of resistance mechanisms through gatekeeper mutations and compensatory pathways, and narrow therapeutic windows due to essential mitotic functions (Jing & Chen, 2021; Bavetsias & Linardopoulos, 2015). These challenges, combined with the absence of reliable biomarkers for patient stratification (Carpinelli & Moll, 2008), have resulted in repeated clinical failures despite their clear oncogenic role. The convergence of abundant structural data in the Protein Data Bank (Berman et al., 2000), extensive datasets from clinical trials (de Groot et al., 2015), and well-characterized structure-activity relationships (Ashraf et al., 2021) make Aurora kinases an ideal testbed for computational design approaches. Additionally, the clear categorization of kinase inhibitor binding modes provides a rational framework for systematic exploration of novel chemical space (Vats et al., 2025).

Protein kinases and Aurora kinases. Protein kinases are regulatory enzymes that covalently transfer a γ-phosphate group from ATP onto a residue of a target protein. Biochemically, protein kinases catalyse the following reaction:

$$\mathrm{MgATP}^{1-} + \text{protein–O:H} \longrightarrow \text{protein–O:PO}_3^{2-} + \mathrm{MgADP} + \mathrm{H}^+,$$

commonly known as phosphorylation. Protein phosphatases are enzymes that catalyze the opposite reaction to phosphorylation, called dephosphorylation. Phosphorylation is thus a reversible protein modification that is regulated by a balance of kinase and phosphatase activity. The addition of a phosphate group is a powerful mechanism that mediates several cellular pathways through functional modification of the target protein — in the form of activation, deactivation, cellular localization and association with other proteins — , especially signalling cascades. Protein kinases can thus be broadly classified according to the target protein's amino acid residue they phosphorylate, which is the free hydroxyl group of one among tyrosine (Tyr, Y), serine (Ser, S) and threonine (Thr, T). However, kinases that are able to phosphorylate a serine residue are also able to do the same for a threonine residue, thus are usually grouped together as serine/threonine kinases. Kinases that phosphorylate tyrosine form another large group identified as threonine kinases. Serine/threonine kinases and tyrosine kinases form the two largest groups, but other minor groups consisting of kinases acting on different amino acid residues exist (e.g. histidine kinases). The deregulation of protein kinase function plays an important role in cancer as well as immunological, inflammatory, neurodegenerative, metabolic, cardiovascular and infectious diseases. For these reasons, protein kinases have become one of the top priority clinical targets and, as a consequence, the development of kinase inhibitors is invested on the most in the pharmaceutical industry (Arter et al., 2022). Pharmacological strategies targeting protein kinases with small molecule drugs most frequently propose ATP-competitive inhibitors. Indeed, many protein kinases have been demonstrated to have a crucial role in carcinogenesis and metastases of various types of cancers, their activity often promoting cell proliferation, survival and migration. When constitutively overexpressed or active, protein kinases can be associated with oncogenesis. Several efforts have been directed towards the study of kinases as drug targets, with the intention of unravelling more personalised and less cytotoxic strategies for treating cancer. One of the most undertaken paths in this direction is the discovery of small molecule kinase inhibitors; however, despite being theoretically promising, they prove to be disappointing when it comes to clinical trials, revealing several unpredicted toxicities and side effects. In addition to that, the major obstacle in the way of a successful treatment of cancers with small molecule kinase inhibitors is acquired pharmacological resistance. Kinase inhibition appears to trigger a strong discerning pressure for cells to acquire resistance to chemotherapy through kinase mutations. The majority of the kinase resistance cases fall into the category of acquired resistance[7], which refers to the progression of a tumour that initially responds to treatment and subsequently

[7] As opposed to *de novo* resistance

becomes resistant to it, without variation in the administration of the inhibitor. Interestingly, a single point mutation in a kinase gatekeeper residue is sufficient to trigger drug resistance, because it prevents the drug inhibitor to form hydrogen bonds with deeper hydrophobic regions of the kinase binding pocket which would only be accessible via "permission" of these residues. Overcoming gatekeeper-mutation-induced drug resistance requires meticulous structural fine-tuning of the drug candidate. Several follow-up strategies have been proposed to work around the problem, notably: i) designing inhibitors that can tolerate a number of different amino acid residues at the kinase gatekeeper position; ii) exploit alternative binding sites of the kinase for inhibitor binding; iii) focus on targeting other pathways that may be required for kinase transformation (Bhullar et al., 2018).

Genomic and protein sequence characteristics of Aurora kinases. The Aurora kinase family consists of three highly conserved serine/threonine kinases: Aurora A (*AURKA*), Aurora B (*AURKB*), and Aurora C (*AURKC*). They are encoded by genes respectively located on chromosomes 20q13, 17p13, and 19q13. Aurora A and B are ubiquitously expressed, whereas Aurora C is predominantly found in the testes but can be overexpressed in cancers. The genes encoding the three human paralogue Aurora kinases map to regions of the DNA that are frequently affected by chromosomal abnormalities in different cancer types. A validation of this statement is provided by the recorded overexpression of each of the three human Aurora genes in tumour cell lines (Carmena & Earnshaw, 2003). The chromosomal loci of the three Homo Sapiens genes are A:20q13, B:17p13 and C:19q13 (Meraldi et al., 2004), with genomic sequences available on UCSC Genome Browser (Casper et al., 2025) at loci chr20:56369390-56392215, chr17:8204734-8210575, chr19:57231009-57235417, respectively, according to human genome assembly GRCh38/hg38 of 2013[8].

Structural biology and target characterization of Aurora kinases. The goal of this step is to understand the 3D structure and active/binding sites of the target. Well-established methods include X-ray crystallography, NMR spectroscopy, cryo-EM, and homology modelling (when a reference structure is not available). More recent methods include AlphaFold (Abramson et al., 2024), Chai-1 (team et al., 2024), and Boltz-2 (Passaro et al., 2025). The high structural similarity of Aurora kinases among themselves and with other protein kinases makes for a challenging drug target, raising the relevant problem of drug-target specificity.
Structurally, Aurora kinases share a conserved kinase domain ($\sim 250 - 270$ residues) with high sequence homology ($\sim 70\%$ identity between Aurora A and B), but diverge in their N- and C-terminal regulatory domains. A distinguishing feature is the absence of a SH3 domain and the presence of a unique D-box (destruction box) and A-box, which regulate protein stability and degradation. Aurora kinases adopt a canonical bilobal kinase fold, with an N-terminal lobe (N-lobe) rich in β-sheets and a C-terminal lobe (C-lobe) dominated by α-helices. The ATP-binding pocket, located at the interlobe cleft, contains conserved residues (e.g., Lys162 in Aurora A, Lys106 in Aurora B) critical for catalytic activity (Carmena et al., 2015). Key structural motifs include:

- Activation loop (T-loop): Phosphorylation of Thr288 (Aurora A) or Thr232 (Aurora B) is essential for full kinase activity.
- P-loop (Gly-X-Gly-X-X-Gly): Anchors ATP via glycine-rich flexibility.
- DFG motif (Asp-Phe-Gly): Regulates Mg^{2+}-ATP coordination; the "DFG-in" conformation is active.
- Aurora kinases are distinguished from other kinases by their small, hydrophobic gatekeeper residue (e.g., Leu215 in Aurora A), which influences inhibitor binding specificity. Unlike many kinases, Aurora A lacks a regulatory SH2 domain and instead relies on TPX2 (targeting protein for Xklp2) for spindle localization and activation.

Subcellular localization. Surprisingly, despite the higher level of structural similarity of the three mammalian Aurora paralogues, they have very distinct subcellular localizations and functions (Carmena & Earnshaw, 2003).

Aurora A, *the polar Aurora*: it is normally associated with the centrosome from the time of centrosome duplication through the mitotic exit, and it is also associated with regions of microtubules that are proximal to centrosomes in mitosis. The defining characteristic of the Aurora A subfamily has been its association with centrosomes and regions of microtubules that are proximal to the centrosome itself, and it associates with the centrosomes that are separating during late S/early G2 phase.

[8] `https://genome-euro.ucsc.edu/index.html`

This localization is dynamic and the protein exchanges continuously with the cytoplasmic pool. The association with the centrosome is directed independently by both the amino-terminal region and the carboxy-terminal catalytic domain, but it does not require kinase activity. The protein TPX2 (targeting protein for XKLP2), which has been implicated in Aurora A activation, is required for the localization of the kinase to spindle microtubules, but not to spindle poles.

Aurora B, *the equatorial Aurora*: it forms a complex with two other proteins, called inner centromere protein (INCENP) and survivin, and behaves as a chromosomal passenger protein. Passenger proteins normally associate with centromeric heterochromatin early in mitosis, transfer to the central spindle in anaphase and are amongst the first proteins to localize at the cell cortex where the contractile ring subsequently forms. Chromosomal passenger proteins remain associated with the midbody during cytokinesis. The dramatic movements of passenger proteins during mitosis led to the proposal that they might have a role in the coordination of chromosomal and cytoskeletal events during the cell cycle. Human Aurora B was first identified in a polymerase chain reaction screen for kinases that were overexpressed in tumours. Aurora B expression and activity in proliferating tissues are cell-cycle regulated: expression peaks at the G2-M transition, and kinase activity is maximal during mitosis. More recent studies showed that the association of the kinase with centromeres during metaphase is dynamic: the protein exchanges continuously with the surrounding cytoplasmic pool just like Aurora A. Once the kinase associates with central spindle microtubules during anaphase (which requires kinase activity), its mobility is highly reduced. A subpopulation of Aurora B also seems to be transported by astral microtubules to the equatorial cell cortex.

Aurora C: much less is known about Aurora kinase C. They are specifically expressed at high levels in the testis and show centrosomal localization from anaphase to telophase.

Function. Aurora A and B share significant sequence similarity, particularly within their kinase domains; however, each kinase exhibits unique precise temporal and spatial control by dynamic association with accessory proteins. Both Aurora A and B are mitotic regulators that are often found to be aberrantly expressed in tumour cells. Their activity is dependent on a number of cofactors; both demonstrate increase in kinase activity during mitosis as a result of association with accessory proteins that generate fully competent kinase complexes and by autophosphorylation of critical residues (T288 AurA and T232 AurB) in the kinase activation (T) loop (Kwiatkowski et al., 2012). The gene *AURKA* is ubiquitously expressed and regulates cell cycle events occurring from late S phase though the M-phase. Both activity and protein levels of Aurora A increase from late G2 phase through the M phase, with peak activity in prometaphase. The kinase activity of Aurora A is tightly regulated throughout the cell cycle. It is activated through the phosphorylation of T288 (human sequence) on its activation loop. It can be inactivated through dephosphorylation of T288 by protein phosphatase 1 (PP1). Beyond phosphorylation and dephosphorylation, its activity is also regulated by its expression and degradation.

Each Aurora kinase has specific functions during the cell cycle that are concordant with its subcellular localization (Carmena & Earnshaw, 2003).

Aurora A has a role in centrosome maturation and separation: the high frequency of monopolar mitotic figures in certain *Drosophila aurora* mutants indicate a potential role for the kinase in centrosome separation. In the absence of Aurora A, recruitment of several components of the pericentriolar matrix to the centrosome is deficient, and the microtubule mass of spindles is decreased by about 60%. Furthermore, the morphology of the astral microtubule array is also aberrant. This might partly reflect an impaired function of factors that regulate microtubule dynamics, such as *Drosophila* transforming-acidic-coiled-coil protein (D-TACC), and/or Eg5, a kinesin-like protein that is involved in spindle assembly. Both of these proteins are substrates of Aurora A *in vitro*. TPX2 binds Aurora A at the centrosome and targets it to the microtubules proximal to the pole. TPX2 also regulates the kinase activity of Aurora A, both by counteracting the activity of protein phosphatase PP1 and stimulating Aurora A autophosphorylation at Thr295—the residue in the activation loop of Aurora A that is essential for kinase activity.

Aurora B has a role in chromosome biorientation: after nuclear envelope breakdown, prometaphase chromosomes rapidly establish attachments to a nearby spindle pole, destabilising syntelic attachments of sister chromatids. This might be especially important in *S. cerevisiae*, in which chromosomes are attached to nuclear microtubules for most of the cell cycle and replicated sister kinetochores enter mitosis attached to the same spindle-pole body. How the kinase recognizes syntelic

attachments is not clear, but it has been proposed that tension between amphitelically oriented sister kinetochores stretches them apart enough to separate microtubule-binding sites from Aurora B that is sequestered in the inner centromere, which thereby limits the accessibility of the kinase to its substrate. Aurora B also seems to have an important role in regulating kinetochore-microtubule interactions in higher eukaryotes. Interference with its function by RNAi, microinjection of function-blocking antibodies or treatment with small-molecule inhibitors all cause defects in chromosome congression. The mammalian kinetochore-specific histone-H3 variant CENP-A is a substrate of Aurora B in mammalian cells. Phosphorylation of CENP-A by Aurora B peaks in prometaphase. Surprisingly, phosphorylation-site mutants show a delay in the late stages of cytokinesis. Why a kinetochore protein should show defects in completion of cytokinesis, a cytoskeleton/membrane event, is unclear.

Aurora C is still an obscure topic: very little is known about this Aurora kinase, except that it seems possible that it might act similarly to Aurora B.

Role in cancer. Aurora kinases are frequently overexpressed/amplified in cancers (e.g., AURKA in breast, ovarian, and gastric cancers; AURKB in colorectal and hematologic malignancies). Their dysregulation leads to genomic instability, aneuploidy, and chemoresistance. Aurora A's oncogenic role extends beyond mitosis, involving NF-κB and MYC stabilization, while Aurora B overexpression promotes polyploidy and metastatic potential.

Inhibitors. An important step in this computational drug design pipeline is to try to describe the common structural features that underpin protein kinase regulatory mechanisms and how they relate to kinase inhibitor discovery (Arter et al., 2022).
In the hunt for a novel class of anticancer treatments, substantial research is being done on the inhibition of crucial regulatory mitotic kinases by using ATP-competitive small-molecules (Kwiatkowski et al., 2012). The inhibition of critical regulatory mitotic kinases using ATP-competitive small molecules is an active area of research in the quest for a new class of anticancer therapeutics. Numerous compounds targeting key cell cycle kinases including Cyclin-dependent kinases (Cdk), Aurora (Aur), Polo-like kinases (Plk) and the kinesin-5 molecular motor have been advanced into the clinical testing. The clinical rationale for targeting mitosis to treat cancer is provided by Taxol, a highly successful anticancer agent that arrests cell division by stabilizing microtubule polymers and thereby disrupting the cellular machinery required for mitotic spindle assembly. Unfortunately, to date most of the small molecules targeting cell cycle kinases have displayed limited clinical efficacy and have suffered from dose-limiting bone marrow toxicity.
Protein kinases are high-priority targets for drug discovery in oncology and other disease settings, and kinase inhibitors have transformed the outcomes of specific groups of patients. Most kinase inhibitors are ATP competitive, deriving potency by occupying the deep hydrophobic pocket at the heart of the kinase domain. Selectivity of inhibitors depends on exploiting differences between the amino acids that line the ATP site and exploring the surrounding pockets that are present in inactive states of the kinase. More recently, allosteric pockets outside the ATP site are being targeted to achieve high selectivity and to overcome resistance to current therapeutics. It is important to keep in mind what are the key regulatory features of the protein kinase family, describe the different types of kinase inhibitors, and highlight examples where the understanding of kinase regulatory mechanisms has gone hand in hand with the development of inhibitors.
The structural similarity of the protein kinase ATP binding site has allowed the identification of common hinge-binding scaffolds and enabled the prediction of binding modes and key interactions of novel ligands providing a rationale for inhibitor design. Approaches for developing ATP-competitive inhibitors are therefore relatively mature and can be used to rapidly generate libraries of compounds for inhibitor hit discovery.
The gatekeeper residue is a frequent site of point mutations in kinases that confer resistance, and therefore, it is a logical strategy to use this property to finetune kinase inhibitor selectivity. ATP-competitive kinase inhibitors are usually classified into three "types", all of which mimic ATP and make interactions with the hinge region within the kinase ATP site but vary in the conformational state of the kinase they interact with, and the additional interactions they make:

- Type I inhibitors are defined as small molecules that bind to the active conformation of the kinase (DFG-in and C-in).
- Type I½ inhibitors bind to the inactive DFG-in, C-out conformation of the kinase.
- Type II inhibitors bind to the inactive DFG-out conformation of the kinase (C-in or C-out).

While type I inhibitors are mostly confined to the ATP site, type I½ and type II inhibitors extend into the distinctive pockets that are opened up in the specific inactive conformations they bind. Inhibitors that do not compete with ATP provide the opportunity to gain greater selectivity over other kinases by binding to allosteric sites that are less conserved than the ATP-binding site. Allosteric inhibitors that bind adjacent to the ATP site are classified as type III and those that bind to other more distant sites as type IV.
Another important class of kinase inhibitors are covalent kinase inhibitors or type VI inhibitors. These compounds exploit the presence of reactive amino acids in the ATP site, typically cysteine, although other residues such as lysine have also been successfully used. These cysteines are not required for catalysis, but they can be used to irreversibly block the ATP site through attachment of a small molecule (Kwiatkowski et al., 2012).
The members of the Aurora kinase family play critical roles in the regulation of the cell cycle and mitotic spindle assembly and have been intensively investigated as potential targets for a new class of anticancer drugs. We describe a new highly potent and selective class of Aurora kinase inhibitors discovered using a phenotypic cellular screen. Optimized inhibitors display many of the hallmarks of Aurora inhibition including endoreduplication, polyploidy and loss of cell viability in cancer cells. Structure-activity relationships with respect to kinome-wide selectivity and guided by an Aurora B co-crystal structure resulted in the identification of key selectivity determinants and discovery of a subseries with selectivity toward Aurora A.

Pathway integration. Aurora kinases intersect with critical oncogenic pathways: Aurora A activates the PI3K/AKT/mTOR axis and stabilizes N-MYC in neuroblastoma; Aurora B cooperates with the SAC (via BubR1 and Mad2) and the RhoA/ROCK pathway during cytokinesis; Cross-talk with p53: Aurora A phosphorylates p53 at Ser215, inhibiting its tumor-suppressor function (Arter et al., 2022).

The Aurora kinase family represents a nexus of mitotic regulation and oncogenic signaling, with structural and functional nuances that inform targeted drug design. Their conserved yet distinct features—localization, activation mechanisms, and pathway interactions—make them compelling targets for anticancer therapies, particularly in AI-driven drug discovery where predictive modeling of inhibitor binding and resistance is paramount.

C.2 DE NOVO LIGAND DESIGN

C.2.1 BINDING SITE ANALYSIS

The goal of this step is to define the binding pocket of the target in terms of size, shape, hydrophobicity, and key residues. Several tools are available for this task, or alternatively, one can turn to literature when the target is particularly well-documented.
The Aurora kinase family shares a conserved catalytic kinase domain, but subtle differences in their ATP-binding pockets influence inhibitor specificity. The binding pocket is typically divided into five key regions: (1) the hinge region, which forms hydrogen bonds with ATP/inhibitors; (2) the hydrophobic back pocket; (3) the phosphate-binding region (P-loop); (4) the gatekeeper residue; and (5) the DFG motif (Asp-Phe-Gly), which regulates activation.
Aurora A has a relatively shallow and flexible binding pocket, with a small gatekeeper residue (Thr217 in humans) allowing broad inhibitor access. Its hinge region (Glu211-Ala213) often interacts with inhibitors via hydrogen bonds, while the hydrophobic back pocket (e.g., Leu139, Leu194) accommodates aromatic moieties. Notably, Aurora A's P-loop (Gly216-Gly222) is highly dynamic, enabling induced-fit binding of diverse compounds.
Aurora B, in contrast, has a more narrow and rigid pocket due to its larger gatekeeper (Leu223) and a unique C-lobe extension that stabilizes the CPC. The hinge region (Leu227-Ala229) favors hydrophobic interactions, and the back pocket (Leu154, Tyr156) is less accessible, explaining why many Aurora B inhibitors are bulkier. The DFG motif in Aurora B adopts an active conformation more readily, influencing inhibitor binding kinetics.
Aurora C closely resembles Aurora B but has a slightly more open pocket near the glycine-rich loop (due to substitutions like Val147), which can enhance binding affinity for certain pan-inhibitors.
The high homology between Aurora A/B pockets complicates drug design by creating some selectivity challenges. However, Aurora A-selective inhibitors (e.g., MLN8054) exploit its flexible P-loop, while Aurora B-selective compounds (e.g., AZD1152) target its distinct C-lobe interactions.

Recent strategies focus on allosteric inhibitors or covalent binders targeting non-conserved cysteines (e.g., Cys290 in Aurora A) to improve specificity.

C.2.2 CHOICE OF GRANULARITY AND DESIGN STRATEGY

The goal of this step is to decide the granularity of the building blocks we are going to use for our generation process. Available options for this step are fragment-based and atom-based drug design. In fragment-based design the algorithm starts with small chemical fragments (typically < 300 Da) that weakly bind to different parts of the target, then grows, links, or mergse them; in atom-based design the algorithm builds molecules atom-by-atom, choosing where to place each atom based on the environment of the binding pocket. After choosing the granularity of the representation, one can choose the strategy guiding the construction, meaning what algorithmic strategy is used to decide where and how to add or modify molecular components: we choose an AI/ML-based strategy and use graph generative models.

GraphBP. For this task, we choose the 3D equivariant generative model GraphBP (Liu et al., 2022), a flow-based graph generative model that designs 3D molecular structures directly inside protein binding pockets. The model operates autoregressively, adding atoms one at a time in 3D space, conditioned on the growing molecule and the encoded protein context. The pocket and any initial fragment are encoded via a SE(3)-equivariant GNN, producing local and global context embeddings. At each step, the model samples atom types using a classifier, and positions using a conditional radial flow over spherical coordinates relative to a reference atom. Atom coordinates are transformed back to global 3D space using learned local frames. Bonds are inferred post hoc using geometry-based heuristics or cheminformatics tools. In this work, we retrain the model on a subset of the original CrossDocked2020 dataset as Specified in Appendix D.

C.3 VIRTUAL SCREENING AND SCORING

The goal of this step is to screen de novo molecules against the target, then score them based on binding affinity, physicochemical properties, and drug-likeness. Methods commonly used to carry out this step include proxy scores for synthetic accessibility, docking and molecular dynamics simulation tools After generating ligands for the Aurora kinase proteins, we calculate the SA_Score, SC-Score, NP_score, SMILES length, Tanimoto similarity[9] with commercialized inhibitors and within the group of generated ligands and Lipinski's rules violations. We visualize the distributions of the scores and investigate minimum, average and maximum values. We begin our post-hoc filtering process based on the "inter" Tanimoto similarity, and only retain the top scoring 100 unique ligands. The rationale behind this choice stems from the observation that the generative model does not follow a biased generation process, thus simply generates general ligands for the target protein; since we are interested in verifying whether we can use the pipeline to produce real world scenario ATP-competitive Aurora kinase inhibitors, we filter out all those molecules having negligible similarity value with existing inhibitors. Despite this choice being logic, it does not ensure the selection of a set of ligands that are actually promising Aurora kinase inhibitors. The first filtering phase is followed by a second one according to the SA_Score value: many have pointed out that this proxy score is the most reliable among the existing ones to give an idea of the synthetic accessibility of the molecule. We keep the 50 top scoring molecules according to SA_Score among the 100 top scoring molecules previously selected according to Tanimoto similarity with known inhibitors.

EquiBind. EquiBind (Stärk et al., 2022) is a deep learning model that predicts how a ligand fits into a protein's binding site in one step, without traditional docking simulations. This process is called blind docking: it consists in predicting a ligand's binding site and pose on a protein without knowing the binding site beforehand, instead searching the entire protein surface. It uses a graph neural network with geometric awareness to understand both the ligand and the protein. The model outputs the ligand's 3D bound pose directly. It's much faster than classical methods and works without needing prior knowledge of the binding pocket. EquiBind formulates molecular docking as a one-shot prediction problem using a geometric deep learning framework. It leverages a SE(3)-equivariant GNN architecture to process 3D structures of ligands and proteins as graphs. Rather than sampling or optimizing poses, EquiBind directly predicts the ligand's final binding conformation via a single forward pass. The model is trained with a supervised loss that aligns the predicted ligand pose to ground truth using RMSD.

[9] Using `rdkit.DataStructs.cDataStructs.BulkTanimotoSimilarity`

AutoDock Vina. AutoDock Vina (Trott & Olson, 2010) is an open-source molecular docking program designed to predict how small molecules, such as drug candidates, bind to a target protein. It is renowned for its speed and accuracy, employing an improved scoring function and efficient search algorithms to model ligand-receptor interactions. Unlike its predecessor, AutoDock, Vina uses a gradient-based optimization method to explore binding poses, significantly reducing computational time while maintaining high precision. The software supports flexible ligand docking and allows limited protein flexibility, making it useful for studying binding modes in proteins like Aurora kinases. Users can define a search space (grid box) around the target's binding pocket, and Vina evaluates binding affinities (in kcal/mol) to rank potential poses. Its command-line interface and compatibility with tools like PyMOL and ChimeraX streamline workflow integration. While Vina lacks explicit solvent or entropic terms in its scoring function, its balance of speed and reliability has made it a staple in virtual screening and structure-based drug design.

C.4 In silico filtering

The goal of this step is to reduce the number of candidate molecules before synthesis. The prediction of synthesizability is the task of estimating how easily a given molecule might be synthesizable in the laboratory based on its structural complexity (Chen & Jung, 2024). We calculate synthesizability according to five different scores: length of SMILES string, SA_score, SCScore, NP_score, and QED score. Furthermore, we analyze in silico values of drug-likeness by using the Lipinski rule of 5, and notice that violations of the rules occur quite often, both in commercialized ligands and in generated ligands.

C.5 Statistical analysis

Since we are using different metrics to assess the synthetic accessibility of the generated compounds, we make sure of whether they represent significantly different ways of measuring synthetic accessibility, and what is the behaviour of some of them with respect to the others. We use SciPy's `mannwhitneyu` function to calculate the significance of the difference between the metrics' distributions. The Mann-Whitney U test is a nonparametric test of the null hypothesis that the distribution underlying one sample is the same as the distribution underlying the alternative sample. In all cases, we note that the test signals a very significant difference in distribution between all combinations of the metrics.

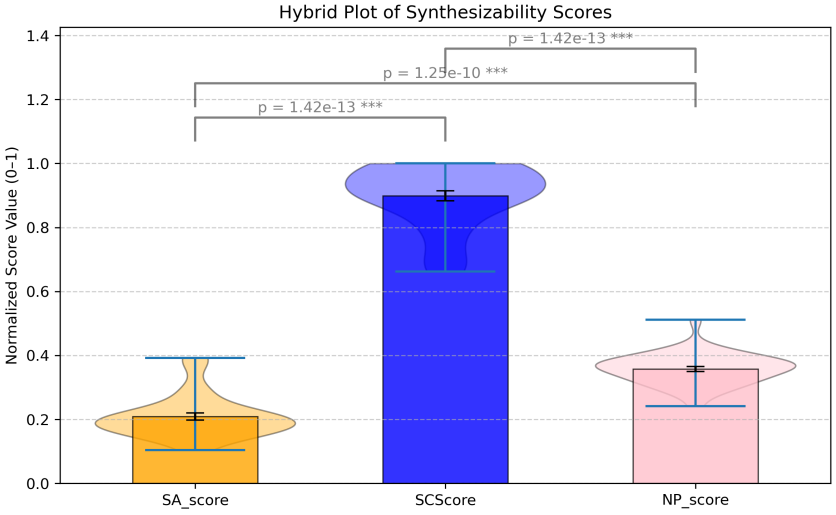


Figure 3: Results of the Mann-Whitney U test on the synthetic accessibility scores distributions.

D EXPERIMENTAL SETTING

D.1 DATA PREPARATION AND QUALITY CONTROL

Before retraining the molecular generative model, we remove from the training data all biological types of interest for current and future analyses. These types include: Aurora kinases, GPCRs, GTPases, Ion channels, and lactamases. We ensure no data leakage between training and testing sets by accurately removing all the data points corresponding to these targets.

D.2 MOLECULAR GENERATIVE MODEL RETRAINING

To ensure generalization capabilities, we retrain our molecular generative models of choice, GraphBP, by using a reduced version of the original CrossDocked2020 dataset, with 1.7% of the dataset removed, and 98.3% of the dataset kept for retraining. We keep off-the-shelf parameters, and subsequently perform inference of the biological types of interest.

D.3 INFORMATION EXTRACTION

Most approved small-molecule drugs have between 20 and 100 atoms (excluding hydrogen atoms). If we include hydrogens, the total atom count usually ranges from 50 to 300 atoms. On the other hand, the smallest approved drugs can have as few as 10–15 non-hydrogen atoms (e.g., lithium (1 atom) in lithium carbonate or ethanol (9 atoms) in some formulations). However, most drugs commonly considered "small molecules" have at least 15–20 non-hydrogen atoms to achieve sufficient potency and selectivity. The upper limit for small-molecule drugs is around 100–150 non-hydrogen atoms (or about 500 atoms total). Beyond this, compounds start entering the realm of "medium-sized molecules" (e.g., macrocycles, peptides). We use this information to condition the model in two of our experiments to retrict the generation process to a pre-defined range of atom numbers.

D.4 EVALUATION

D.4.1 POST-HOC EVALUATION

SMILES length. SMILES length is a very simple heuristic that associates molecules with longer SMILES strings as an indication of synthetic difficulty. The length of a SMILES string correlates closely with the number of heavy atoms in a molecule (i.e., larger molecules are harder to synthesize) but is further increased by the presence of formal charges, ring closures, and defined stereochemistry. The SMILES length is an indicator of the complexity of a chemical compound and can be used as a predictor of synthetic accessibility (Gao & Coley, 2020). Generally, the longer a SMILES, the harder it is to synthesize a molecule. We calculate the direct length of the molecular SMILES using the Python's built-in `len()` function.

Tanimoto similarity. The concept of chemical similarity between two chemical elements, compounds, or molecules refers to their similarity in terms of either structural or functional qualities, meaning the effect that that chemical entity has on other reaction partners in inorganic or biological settings. The various existing metrics of similarity calculation rely on a fundamental principle defined by Johnson and Maggiora[10] , which states that "similar compounds have similar properties". The most popular similarity measure for comparing chemical structures represented with fingerprints is indeed the Tanimoto (or Jaccard) coefficient T which is obtained by calculating the intersection over union of two sample sets:

$$T(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

By design, $0 \leq T(A, B) \leq 1$, with $T(A, B) = 0$ meaning no similarity and $T(A, B) = 1$ meaning equality.

Chemical fingerprints. Chemical fingerprints are binary vector representations of molecular structures that encode the presence or absence of specific structural features, such as substructures, func-

[10]Johnson, A.M. and Maggiora, G.M. (1990). Concepts and Applications of Molecular Similarity. New York: John Wiley & Sons.

tional groups, or topological patterns. Common fingerprint types include ECFP (Extended Connectivity Fingerprints, also known as Morgan fingerprints), MACCS keys, and topological fingerprints. These bit vectors enable rapid comparison of molecular similarity through the Tanimoto coefficient, calculated as the ratio of shared bits to total bits: $T = (A \cap B)/(A \cup B)$, where A and B are fingerprint vectors. Tanimoto scores range from 0 (no similarity) to 1 (identical), with values > 0.85 typically indicating high structural similarity. This metric is widely used in virtual screening, clustering chemical libraries, and assessing the novelty of generated molecules.

Synthetic accessibility. We calculate synthesizability according to four different scores: length of SMILES string, SA_score, SCScore, and NP_score.
SA_Score (Ertl & Schuffenhauer, 2009) is computed as a weighted combination of fragment contributions and molecular complexity penalties. It ranges from 1 (simple) to 10 (complex).
SCScore (Coley et al., 2018) quantifies synthetic complexity as a learned function correlating with the expected number of reaction steps required to synthesize a molecule, implicitly accounting for starting material availability. It ranges from 1 (simple) to 5 (complex).
NPscore (Ertl et al., 2008) quantifies the "natural-product-likeness" of a compound. It ranges from -5 (simple) to 5 (complex).

Drug-likeness. Quantitative Estimate of Drug-likeness (QED) (Bickerton et al., 2012) is a widely used metric that provides a numerical assessment of how closely a molecule resembles known oral drugs. QED combines eight molecular properties—molecular weight, lipophilicity (logP), hydrogen bond donors and acceptors, polar surface area, rotatable bonds, aromatic rings, and structural alerts—into a single score ranging from 0 to 1, where higher values indicate greater drug-likeness. Unlike binary rule-based filters such as Lipinski's Rule of Five, QED provides a continuous, quantitative measure that reflects the desirability of each property based on distributions observed in approved oral drugs.

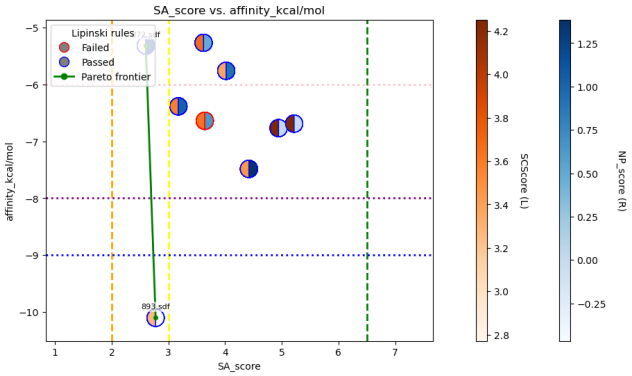
Lipinski's rule of 5. Lipinski's Rule of Five (Lipinski et al., 2001) and Veber's rules are foundational guidelines for assessing oral drug-likeness based on binary thresholds. Lipinski's rules state that oral drugs typically have molecular weight ≤ 500 Da, logP ≤ 5, hydrogen bond donors ≤ 5, and acceptors ≤ 10. Veber's rules focus on molecular flexibility and polarity, requiring ≤ 10 rotatable bonds and polar surface area ≤ 140. These rules provide simple pass/fail criteria but don't quantify how "drug-like" a molecule is. In contrast, QED offers a continuous score (0-1) by integrating similar properties with weighted desirability functions based on distributions in approved drugs.

Binding affinity. Binding affinity quantifies the strength of interaction between a ligand and its target protein, typically expressed as a dissociation constant (Kd) or inhibition constant (Ki). In computational drug discovery, binding affinity is often reported in kcal/mol as the Gibbs free energy of binding (ΔG), which relates to Kd through the equation $\Delta G = RT \ln(Kd)$, where R is the gas constant and T is temperature. More negative ΔG values indicate stronger binding; for example, ΔG = -10 kcal/mol corresponds to nanomolar affinity (Kd $\approx$ 10-50 nM), while ΔG = -7 kcal/mol indicates micromolar binding (Kd $\approx$ 1-10 μM). Accurate prediction of binding affinity remains challenging due to entropic effects, solvation, and conformational changes, with docking scores often correlating poorly with experimental values.

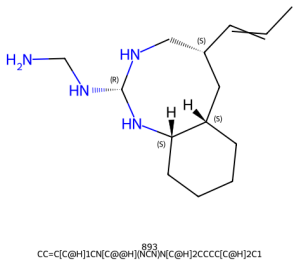
E ADDITIONAL RESULTS

E.1 AURORA KINASE B EXPERIMENTS

E.1.1 UNCONSTRAINED, UNBIASED, BINDING POCKET UNCONDITIONAL GENERATION



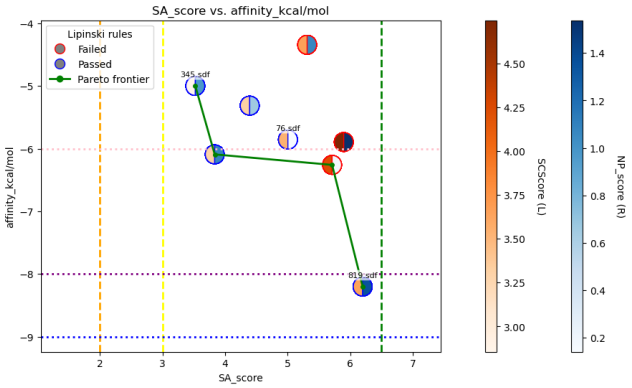
(a) Pareto frontier of Aurora kinase B ligands generated with the first experiment. The plot shows the molecules having passed all the stages and reports synthetic accessibility scores, binding affinity, and Lipinski's rules compliance.



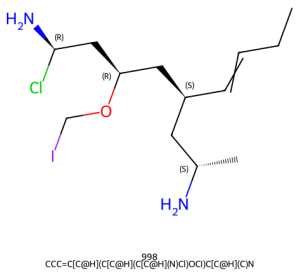
(b) Example of one of the molecules in the Pareto frontier. The visualization has been obtained with RDKit.

Figure 4: Results of the first experiment after benchmarking. The model does not use prior information, is not biased with any synthetic accessibility score, and is not equipped with the target's binding site information.

E.1.2 UNCONSTRAINED, UNBIASED, BINDING POCKET CONDITIONAL GENERATION



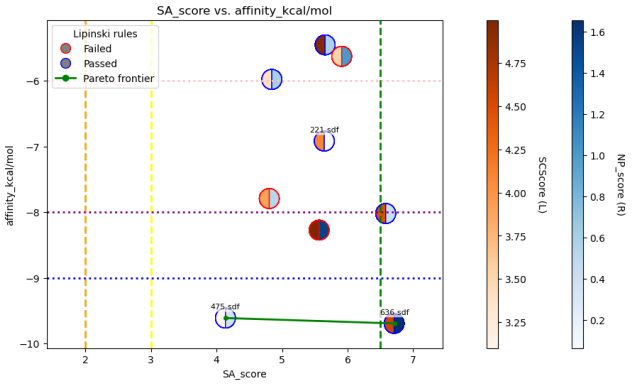
(a) Pareto frontier of Aurora kinase B ligands generated with the second experiment. The plot shows the molecules having passed all the stages and reports synthetic accessibility scores, binding affinity, and Lipinski's rules compliance.



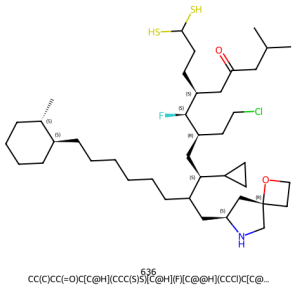
(b) Example of one of the molecules in the Pareto frontier. The visualization has been obtained with RDKit.

Figure 5: Results of the second experiment after benchmarking. The model does not use prior information, is not biased with any synthetic accessibility score, but is now equipped with the target's binding site information.

E.1.3 CONSTRAINED, UNBIASED, CONDITIONAL GENERATION



(a) Pareto frontier of Aurora kinase B ligands generated with the second experiment. The plot shows the molecules having passed all the stages and reports synthetic accessibility scores, binding affinity, and Lipinski's rules compliance.



(b) Example of one of the molecules in the Pareto frontier. The visualization has been obtained with RDKit.

Figure 6: Results of the final experiment after benchmarking. The model now uses prior information about the desired number of atoms of the ligands to generate, is not biased with any synthetic accessibility score, but is now equipped with the target's binding site information.

F IMPLEMENTATION DETAILS

The project serves as a ready-to-use toolbox implemented using PyTorch, RDKit, Snakemake and GitHub submodules for modularity.
Among the benefits of structuring the pipeline with GitHub submodules and Snakemake workflow management there are:

- Modularity: each component can be developed/tested independently.
- Reusability: post-processing tools can be used in other projects.
- Maintainability: clear separation of concerns.
- Collaboration: team members can work on different components.
- Versioning: each components has its own version control.
- Testing: individual components can be unit tested.

Pipeline-specific code in the main repository:

- Workflow orchestration as Snakemake rules.
- Configuration management as config files.
- Cross-component analysis as comparison of results across steps.
- Visualization and reporting as pipeline-wide plots and reports.
- Integration tests for testing the complete pipeline.

This structure maintains a cohesive pipeline workflow while providing the user with modular and reusable components.
Currently supported submodules are:

- GraphBP: For ligand generation; supported receptors: PDB ID 4AF3.
- `hope-box`: For post-hoc filtering; includes synthesizability and similarity analysis.
- EquiBind: For blind docking.
- `vina-box`: AutoDock Vina for molecular dynamics simulations.