A PROSPECTIVE EXPERIMENTAL ASSESSMENT OF STRUCTURE-BASED DRUG DESIGN: ARE WE THERE YET?

Maximilian G. Schuh[1*] **Joshua Hesse**[1*] **Konstantin Eckel**[1*] **Văn Quý Công**[2]

Niklas Halbwedl[2] **Frederik Weber**[3] **Andres Jäschke**[3]

Martin Zacharias[2] **Stephan A. Sieber**[1†]

[1] Chair of Organic Chemistry II, TUM School of Natural Sciences, Technical University of Munich
[2] Chair of Theoretical Biophysics, TUM School of Natural Sciences, Technical University of Munich
[3] Institute of Pharmacy and Molecular Biotechnology, Heidelberg University
[*] These authors contributed equally. The order was determined by a coin flip.
[†] stephan.sieber@tum.de

ABSTRACT

Artificial intelligence (AI) is reshaping drug discovery, particularly through structure-based drug design (SBDD), which promises pocket-aware generation of ligands tailored to a chosen target. Yet, most evidence for the success of generative structure-based drug design comes from retrospective benchmarks and curated case studies, often without compound realization and experimental validation—leaving the question open whether these models generalize to truly novel targets. This gap is especially relevant in antibiotic discovery, where data is sparse and true hits are rare. The key question of this paper is whether off-the-shelf structure-aware generators can prospectively deliver active molecules for previously unseen antibacterial targets. Here, we show that a standardized end-to-end evaluation of six state-of-the-art SBDD models yields extremely few experimentally supported hits. Across three structurally selected bacterial targets, >100 000 generated designs were reduced to a small set after standardized post-processing, medicinal-chemistry filtering, and commercial analogue mapping. For the most promising target, MurC, 47 prioritized compounds were acquired and tested, and only one showed weak measurable inhibition at the screening concentration. Together, these results suggest a substantial gap between plausible-looking, protein-conditioned designs and reproducible biochemical activity under realistic discovery constraints, underscoring the need for prospective, assay-closed benchmarks to guide the next generation of SBDD methods.

1 INTRODUCTION

The integration of AI into drug discovery has profoundly reshaped how researchers approach the development of new therapeutics (Vamathevan et al., 2019; Catacutan et al., 2024; Zhang et al., 2025). In particular, recent advances in deep learning (DL)-based molecular generation have raised expectations that *de novo* design of drug-like compounds could transition from a vision to a practical, scalable reality (Grisoni, 2023; Schneider & Clark, 2019). However, despite impressive methodological progress, a critical question remains unanswered: do current AI-driven, SBDD approaches work in practice when applied prospectively to novel targets and evaluated under realistic experimental constraints?

In this work, we set out to directly address this gap through a prospective evaluation of AI-driven SBDD. Our central aim is not to optimize a single model or target, but to benchmark whether state-of-the-art generative methods can design viable ligands for novel, previously unseen targets. To this

end, we first identified antibacterial targets which partially lie outside the training distributions of the evaluated models, thereby probing their true design and generalization capabilities rather than their ability to interpolate over familiar chemotypes or binding modes.

We systematically evaluate six contemporary generative models spanning diffusion-based, autoregressive, graph neural network (GNN), and language-model-based architectures (Li et al., 2024; Zhang et al., 2023; Schneuing et al., 2024; Peng et al., 2022; Guan et al., 2023; Wu et al., 2024). All models are applied in a standardized workflow, and their outputs are subjected to identical post-processing, filtering, and prioritization procedures to ensure comparability. Importantly, we treat compound realization as a core component of the benchmark: generated molecules are curated for chemical validity and developability, mapped to commercially available analogues where possible (Walters et al., 1998; 1999).

Ultimately, selected compounds were evaluated experimentally for target binding and antibacterial activity, closing the loop from *in silico* design to *in vitro* validation. By advancing AI-generated molecules through this complete pipeline, our study provides a prospective, experimentally grounded assessment of the current state of SBDD. Rather than asking whether these models can generate plausible molecules, we ask a more stringent and practically relevant question: whether today's AI-driven design tools can reliably deliver active compounds against novel targets when tested under realistic discovery conditions.

2 METHODS

Six pretrained, publicly released SBDD generators were evaluated to span major model classes: DeepBlock, DiffSBDD, Pocket2Mol, ResGen, TamGen, and TargetDiff (Li et al., 2024; Schneuing et al., 2024; Peng et al., 2022; Zhang et al., 2023; Wu et al., 2024; Guan et al., 2023). All methods were executed from the authors' repositories using the provided weights.

Target discovery and selection Predicted proteomes from the AlphaFold Database were collected for seven clinically relevant pathogens (*E. coli, E. faecium, S. aureus, K. pneumoniae, P. aeruginosa, H. pylori,* and *M. tuberculosis*) and clustered with Foldseek to identify conserved structural protein families (van Kempen et al., 2024). Clusters were prioritized by enrichment in essential genes across pathogens and filtered against a *Homo sapiens* structure set to remove human analogs.

Protein inputs and binding-site definition Target structures were sourced from AlphaFold predictions; for CdsA, AlphaFold 3 complex predictions with the native ligand was used for pocket localization. Binding-site residues were defined from UniProt annotations (MurC), substrate-associated pocket definition (CdsA), or an AlphaFold-predicted dimerization interface (CohE).

Generation and standardized curation For each target–model pair, ligands were sampled (up to 10 000 per model) with multiple random seeds when applicable. All outputs were harmonized via the same post-processing workflow: simplified molecular-input line-entry system (SMILES) parsing/deduplication, ChEMBL-based standardization, and removal of unstable rings and rapid elimination of swill (REOS) (Dundee) alert annotations (Walters et al., 1999; Walters, 2024). Prioritization applied antibacterial-relevant physicochemical constraints (molecular weight; calculated logD at pH 7.4), then mapped commercial analogs with SpaceMACS across make-on-demand (maximum common substructure (MCS) similarity ≥ 0.9) (Schmidt et al., 2022). Ligand–target pairs were further filtered using AlphaFold 3 complex predictions, retaining candidates predicted near the predefined pocket center or interface region. For all complexes, molecular mechanics generalized Born surface area (MM/GBSA) energies and ligand root mean square deviation (RMSD) (after protein alignment) were computed per trajectory (Miller et al., 2012), and ligands were required to meet RMSD <3 Å in molecular dynamics (MD) simulations. Additionally, Boltz 2 with MM/GBSA was used to further refine molecule selection (Passaro et al., 2025).

Affinity and binding experiments Recombinant MurC was purified by His-tag IMAC. Compounds were assessed in parallel using a target-specific *in vitro* activity assay to quantify functional inhibition and an orthogonal biophysical binding assay to test for direct ligand–protein interactions, with replicate measurements compared against matched vehicle controls.

All methods are further detailed in Appendix A.

3 RESULTS

To create a realistic research scenario for SBDD, we chose the identification of a novel antibiotic as our application. The workflow is shown in Figure 1. We first used cross-species structural proteomic clustering over 7 pathogenic bacterial strains of high clinical relevance, using their AlphaFold database proteomes (Varadi et al., 2024) in combination with Foldseek (van Kempen et al., 2024) to identify clusters of structurally conserved essential genes (Appendix A.1). Out of these clusters, we selected 3 proteins: MurC, CdsA, and CohE. These exhibit varying degrees of similarity to the CrossDocked2020 dataset (Francoeur et al., 2020), on which all SBDD methods used in this study were trained on (Figure B1), while offering new and unique antibiotic modes of action. Ligands were then designed, filtered, and, for MurC, experimentally validated.

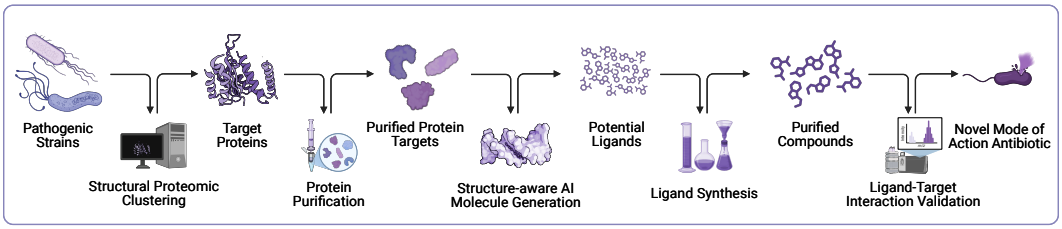

Figure 1: **AI-augmented drug development pipeline.** Seven pathogenic strain proteomes were clustered using Foldseek. Targets were purified and ligands were designed using six structure-aware molecular generative models. Generated structures were rigorously filtered and experimentally tested with the goal to identify compounds with new antibiotic modes of action.

3.1 MOLECULE GENERATION

The following state-of-the-art models were evaluated: DeepBlock, DiffSBDD, Pocket2Mol, ResGen, TamGen, and TargetDiff (Li et al., 2024; Guan et al., 2023; Peng et al., 2022; Zhang et al., 2023; Wu et al., 2024; Schneuing et al., 2024). We aimed to generate 10 000 molecules per target and model. However, not all models achieved this mark, particularly Pocket2Mol and ResGen (Figure B2a). If a model exceeds a certain threshold of failed attempts, it exits the generation loop. The likely cause of not reaching the desired number of designs is poor generalization, which causes the model to get stuck on certain motifs and, in turn, failing to explore a broader chemical space. TamGen could only generate 1000 molecules, so as suggested for LLMs, we used 10 seeds to generate the desired number of structures. Overall, we generated 117 231 molecules across all methods.

Next, we computed the individual percentage of valid molecules generated by each model individually using our `SMILESCleaner` (Appendix A.3), which verifies the validity of SMILES and checks if the ring systems are considered unstable. This step was necessary because internal cleaning methods vary across models, depending on their implementation and architecture. DeepBlock outperformed all other models, with ∼90 % of its generated molecules being valid. Most other models achieved around 50 % validity, except ResGen, which produced only ∼30 % valid molecules (Figure B2b). In total, out of the original 117 231 molecules, 69 867 were found to be valid across all models.

We also analyzed the chemical space and diversity of the generated molecules (Figures B3 and B4). TamGen probed the chemical space the most, while DeepBlock and Pocket2Mol balanced between exploration and exploitation. In contrast, DiffSBDD, ResGen, and TargetDiff produced structurally less diverse molecules compared to other methods and showed pronounced clusters.

3.2 CURATION

Our goal is to efficiently identify promising scaffolds among 69 867 candidates. To narrow the search, we applied filters commonly used in medicinal chemistry and antibacterial drug discovery. These include the REOS/Dundee filters, which attempt to enrich compounds with desirable or “drug-like” properties (Walters et al., 1998; 1999).

First, we analyzed the number of Dundee alerts generated in relation to both the model and the target (Figure B5), a subset of the REOS/Dundee structural alert library. These libraries are derived from medicinal chemistry rule sets to filter for compounds with high drug-likeness. The models DeepBlock and TamGen generated more molecules with fewer REOS/Dundee alerts compared to all other methods. We further investigated the underlying causes of these alerts (Figure B6). The main alert-contributors for MurC were phosphate-containing compounds, which is reasonable as both of MurC's native ligands contain phosphates. Similarly, the most prominent REOS/Dundee alert for CdsA were the presence of aliphatic long chains and phosphates. As this protein is a membrane protein involved in phospholipid and glykolipid synthesis, these results hint to the fact that the biochemical context of these proteins is taken into account during generation.

3.3 Commercial Availability

Chemical vendors provide a vast chemical space as hypothetical compounds, which can be synthesized through automated and fragment-based methods (Table A2). We used SpaceMACS to identify our compounds or close analogs in these vast chemical spaces (Schmidt et al., 2022) and filtered out all compounds with no commercial analogs with MCS similarities of 0.9 or higher.

3.4 Antibiotic Properties

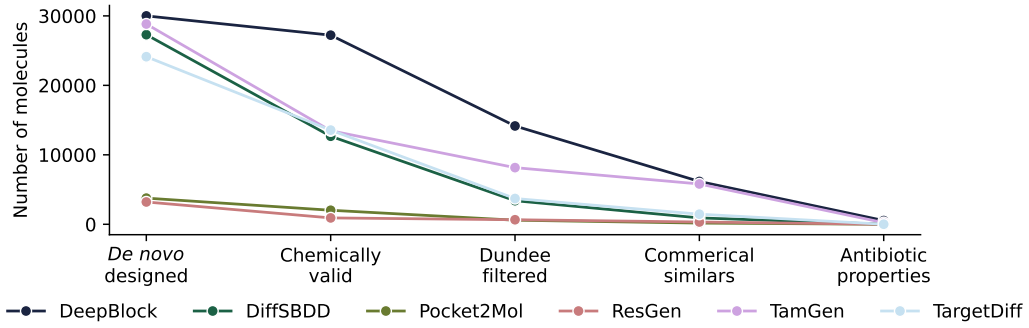

Figure 2: **Processing of *de novo* structures.** We evaluated all models according to the curation and filtering procedure. We checked the validity, removed any molecules that triggered REOS/Dundee alerts, assessed the commercial availability, and applied an additional physicochemical filters associated with antibiotic activity. Shown here are the absolute number of molecules remaining at each step per model.

We then further refined the desired output chemical space using physiochemical principles specific to antibiotic design. Unlike other drug classes, antibiotics exhibit distinct property profiles, allowing us to tailor our search for novel antibacterial scaffolds.(Reck et al., 2019; Brown et al., 2014; Tommasi et al., 2015). Accordingly, we selected compounds with clogD values between −2 to 2 and molecular weights from 200 Da to 450 Da. Figure B7 illustrates this narrowed selection space (target and alert-based investigations can be found in Figure B8). One key observation is that Pocket2Mol, ResGen, and TargetDiff produced few compounds within the desired property window. Since all methods were trained on the same dataset, this discrepancy likely stems from differences in model architecture or training modes. This final filtering step left us with only 814 compounds, representing ∼0.7 % of the original 117 231 designs. The step-by-step elimination of molecules is shown in Figure 2 and Figure B9.

3.5 Affinity and Binding Experiments

We further refined our selection using AlphaFold 3 complex prediction to ensure the ligands were placed in the desired position (Figure B10) and applied MM/GBSA/Boltz 2 ranking to produce the final candidate set (Abramson et al., 2024; Passaro et al., 2025; Miller et al., 2012). These two orthogonal approaches did not correlate, as can be seen in Figure B11. Out of the three selected targets, MurC is the easiest to target as it is the closest to the training set distribution (Figure B1) with

validated experimental workflows and inhibitors available. Therefore, 50 non-redundant candidates for MurC were selected; 44 were synthesizable and purchased. We also included three MurC designs containing phosphates, which were initially filtered out by REOS/Dundee alerts but were retained based on target-specific considerations (see Appendix B.3). A total of 47 compounds (Figure B12) were screened for activity at 100 µM using a malachite green assay. Surprisingly, only one compound showed measurable but very weak inhibition (C_{20}). Then, we performed a thermal shift assay to assess target binding. This assay showed an atypical destabilizing shift for C_{20}, as opposed to the stabilization observed for the known MurC inhibitor, AMP-PCP. (Figure B13). Further, neither affinity nor binding results did correlate with the Boltz 2 and MM/GBSA calculations as no method highlighted C_{20} as more promising than other candidates (Figure B11).

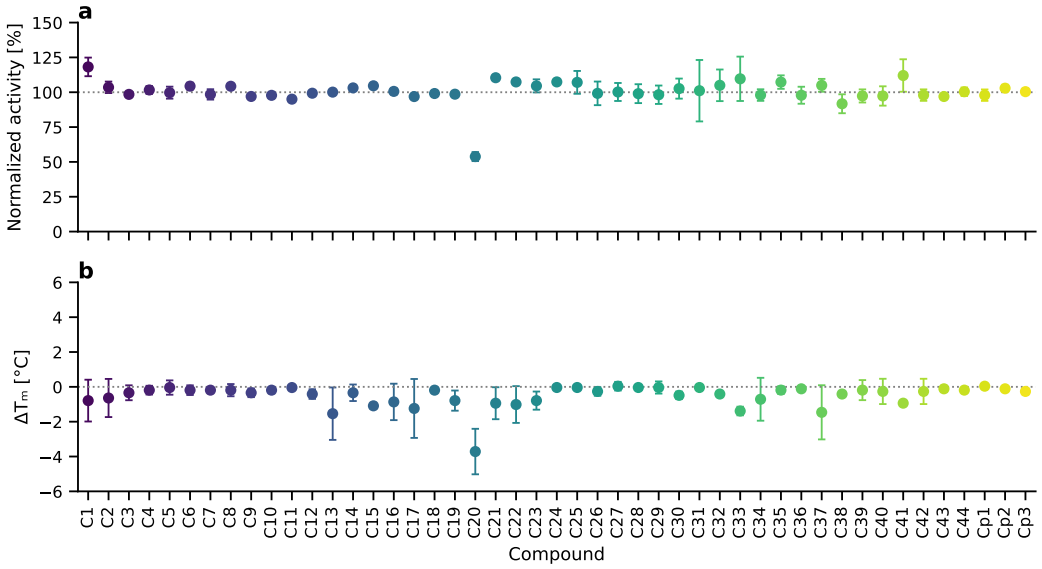

Figure 3: **Experimental validation of designed compounds.** Compound binding and inhibition of MurC were assessed using a malachite green assay (**a**) and thermal shift assay (**b**). Malachite green assay was performed with 3 independent replicates from 3 experiments at 100 µM and normalized to on-plate DMSO controls. Thermal shift assay was performed in 4 technical replicates at 200 µM and shifts calculated to on-plate DMSO controls. Compounds Cp1–3 contain phosphate groups.

4 Conclusion

This study provides a prospective assessment of current SBDD models and highlights a substantial gap between protein-conditioned molecular designs and experimentally confirmed biochemical activity. Although models produced large design pools, rigorous chemical and medicinal-chemistry filtering collapsed these to a small number of synthetically accessible candidates. Strong differences in chemical validity, structural diversity, and drug-likeness emerged between models despite shared training data, suggesting that architectural choices, rather than dataset overlap, drive generalization performance.

Experimentally, the prospective hit rate was low, despite MurC being the most promising target. One out of 47 MurC candidates showed measurable inhibition together with evidence of target engagement, corresponding to a hit rate of ∼2 %. While this exceeds typical confirmed hit rates as low as 0.001 % reported for high-throughput screening (Weissman & Leadlay, 2005), it remains far below the success suggested by many curated SBDD case studies and underscores that ready-to-use workflows still yield few functional hits under realistic constraints. Notably, neither MM/GBSA nor Boltz 2 affinity scores prioritized the weakly active compound, and we observed no meaningful correspondence between these rankings and experimental outcomes, indicating limited prospective utility of common scoring surrogates in this regime. As SBDD failed to generate hits for the most promising target MurC, further target investigations appeared premature with the selected models.

We hypothesize that the evaluated models and scoring schemes do not capture the conformational flexibility of proteins, e.g., at the active site. This mismatch could result in ligands designed for a rarely present or inactive conformational state, resulting in the low hit rate. C_{20} exhibited a destabilizing thermal-shift profile despite inhibition, which suggests an effect on the protein's conformational stability. Overall, these results suggest that the field of SBDD is not yet ready for application to real-world drug discovery settings. However, the field is highly dynamic and new methods are being published frequently, which we will continue to evaluate as time goes on.

ACKNOWLEDGMENTS

The authors thank Merck KGaA Darmstadt for their generous support with the Merck Future Insight Prize 2020. This project is also cofunded by the European Union (ERC, breakingBAC, 101096911).

We would also like to thank Prof. Thomas Schlichthärle and Prof. Martin Steinegger for the many fruitful discussions and suggestions. Further, we thank Aleksandra Daniluk and Anna M. Pietschmann for their valuable feedback and proofreading on the manuscript.

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(8016):493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w.

H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, October 1984. ISSN 0021-9606. doi: 10.1063/1.448118.

Dean G. Brown, Tricia L. May-Dracka, Moriah M. Gagnon, and Ruben Tommasi. Trends and Exceptions of Physical Properties on Antibacterial Activity for Gram-Positive and Gram-Negative Pathogens. *Journal of Medicinal Chemistry*, 57(23):10144–10161, December 2014. ISSN 0022-2623. doi: 10.1021/jm501552x.

David A. Case, Hasan Metin Aktulga, Kellon Belfon, David S. Cerutti, G. Andrés Cisneros, Vinícius Wilian D. Cruzeiro, Negin Forouzesh, Timothy J. Giese, Andreas W. Götz, Holger Gohlke, Saeed Izadi, Koushik Kasavajhala, Mehmet C. Kaymak, Edward King, Tom Kurtzman, Tai-Sung Lee, Pengfei Li, Jian Liu, Tyler Luchko, Ray Luo, Madushanka Manathunga, Matias R. Machado, Hai Minh Nguyen, Kurt A. O'Hearn, Alexey V. Onufriev, Feng Pan, Sergio Pantano, Ruxi Qi, Ali Rahnamoun, Ali Risheh, Stephan Schott-Verdugo, Akhil Shajan, Jason Swails, Junmei Wang, Haixin Wei, Xiongwu Wu, Yongxian Wu, Shi Zhang, Shiji Zhao, Qiang Zhu, Thomas E. III Cheatham, Daniel R. Roe, Adrian Roitberg, Carlos Simmerling, Darrin M. York, Maria C. Nagan, and Kenneth M. Jr. Merz. AmberTools. *Journal of Chemical Information and Modeling*, 63(20): 6183–6191, October 2023. ISSN 1549-9596. doi: 10.1021/acs.jcim.3c01153.

Denise B. Catacutan, Jeremie Alexander, Autumn Arnold, and Jonathan M. Stokes. Machine learning in preclinical drug discovery. *Nature Chemical Biology*, 20(8):960–973, August 2024. ISSN 1552-4469. doi: 10.1038/s41589-024-01679-1.

Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993. ISSN 0021-9606. doi: 10.1063/1.464397.

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, September 2020. ISSN 1549-9596. doi: 10.1021/acs.jcim.0c00411.

Samuel Genheden and Ulf Ryde. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*, 10(5):449–461, May 2015. ISSN 1746-0441. doi: 10.1517/17460441.2015.1032936.

Francesca Grisoni. Chemical language models for de novo drug design: Challenges and opportunities. *Current Opinion in Structural Biology*, 79:102527, April 2023. ISSN 0959-440X. doi: 10.1016/j.sbi.2023.102527.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction, March 2023.

Shahul Hameed P, Praveena Manjrekar, Murugan Chinnapattu, Vaishali Humnabadkar, Gajanan Shanbhag, Chaitanyakumar Kedari, Naina Vinay Mudugal, Anisha Ambady, Boudewijn L.M. de Jonge, Claire Sadler, Beena Paul, Shubha Sriram, Parvinder Kaur, Supreeth Guptha, Anandkumar Raichurkar, Paul Fleming, Charles J. Eyermann, David C. McKinney, Vasan K. Sambandamurthy, Manoranjan Panda, and Sudha Ravishankar. Pyrazolopyrimidines Establish MurC as a Vulnerable Target in Pseudomonas aeruginosa and Escherichia coli. *ACS Chemical Biology*, 9(10):2274–2282, October 2014. ISSN 1554-8929. doi: 10.1021/cb500360c.

Pengyong Li, Kaihao Zhang, Tianxiao Liu, Ruiqiang Lu, Yangyang Chen, Xiaojun Yao, Lin Gao, and Xiangxiang Zeng. A deep learning approach for rational ligand generation with toxicity control via reactive building blocks. *Nature Computational Science*, pp. 1–14, November 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00718-0.

Richard J. Loncharich, Bernard R. Brooks, and Richard W. Pastor. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers*, 32(5):523–535, 1992. ISSN 1097-0282. doi: 10.1002/bip.360320508.

James A. Maier, Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, August 2015. ISSN 1549-9618. doi: 10.1021/acs.jctc.5b00255.

Jonathon C. O. Mifsud, Spyros Lytras, Michael R. Oliver, Kamilla Toon, Vincenzo A. Costa, Edward C. Holmes, and Joe Grove. Mapping glycoprotein structure reveals Flaviviridae evolutionary history. *Nature*, 633(8030):695–703, September 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07899-8.

Bill R. III Miller, T. Dwight Jr. McGee, Jason M. Swails, Nadine Homeyer, Holger Gohlke, and Adrian E. Roitberg. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation*, 8(9):3314–3321, September 2012. ISSN 1549-9618. doi: 10.1021/ct300418h.

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, June 2025.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2Mol: Efficient Molecular Sampling Based on 3D Protein Pockets, May 2022.

Folkert Reck, Johanna M. Jansen, and Heinz E. Moser. Challenges of antibacterial drug discovery. *Arkivoc*, 2019(4):227–244, July 2019. ISSN 1551-7012. doi: 10.24820/ark.5550190.p010.955.

Jean-Paul Ryckaert, Giovanni Ciccotti, and Herman J. C Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *Journal of Computational Physics*, 23(3):327–341, March 1977. ISSN 0021-9991. doi: 10.1016/0021-9991(77)90098-5.

Robert Schmidt, Raphael Klein, and Matthias Rarey. Maximum Common Substructure Searching in Combinatorial Make-on-Demand Compound Spaces. *Journal of Chemical Information and Modeling*, 62(9):2133–2150, May 2022. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c00640.

Gisbert Schneider and David E. Clark. Automated De Novo Drug Design: Are We Nearly There Yet? *Angewandte Chemie International Edition*, 58(32):10792–10803, 2019. ISSN 1521-3773. doi: 10.1002/anie.201814681.

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*, pp. 1–11, December 2024. ISSN 2662-8457. doi: 10.1038/s43588-024-00737-x.

Ruben Tommasi, Dean G. Brown, Grant K. Walkup, John I. Manchester, and Alita A. Miller. ESKAPEing the labyrinth of antibacterial discovery. *Nature Reviews Drug Discovery*, 14(8): 529–542, August 2015. ISSN 1474-1784. doi: 10.1038/nrd4572.

Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, and Shanrong Zhao. Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6): 463–477, June 2019. ISSN 1474-1784. doi: 10.1038/s41573-019-0024-5.

Michel van Kempen, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 42(2):243–246, February 2024. ISSN 1546-1696. doi: 10.1038/s41587-023-01773-0.

Mihaly Varadi, Damian Bertoni, Paulyna Magana, Urmila Paramval, Ivanna Pidruchna, Malarvizhi Radhakrishnan, Maxim Tsenkov, Sreenath Nair, Milot Mirdita, Jingi Yeo, Oleg Kovalevskiy, Kathryn Tunyasuvunakool, Agata Laydon, Augustin Žídek, Hamish Tomlinson, Dhavanthi Hariharan, Josh Abrahamson, Tim Green, John Jumper, Ewan Birney, Martin Steinegger, Demis Hassabis, and Sameer Velankar. AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, 52(D1):D368–D375, January 2024. ISSN 0305-1048. doi: 10.1093/nar/gkad1011.

John M. Walker (ed.). *The Proteomics Protocols Handbook*. Humana Press, Totowa, NJ, 2005. ISBN 978-1-58829-343-5 978-1-59259-890-8. doi: 10.1385/1592598900.

W Patrick Walters. Generative Molecular Design Isn't As Easy As People Make It Look, May 2024.

W. Patrick Walters, Matthew T Stahl, and Mark A Murcko. Virtual screening—an overview. *Drug Discovery Today*, 3(4):160–178, April 1998. ISSN 1359-6446. doi: 10.1016/S1359-6446(97)01163-X.

W Patrick Walters, Ajay A Murcko, and Mark A Murcko. Recognizing molecules with drug-like properties. *Current Opinion in Chemical Biology*, 3(4):384–387, August 1999. ISSN 1367-5931. doi: 10.1016/S1367-5931(99)80058-1.

Frederik Weber, Nikolas Alexander Motzkus, Leona Brandl, Marvin Möhler, Andrijana Alempijevic, and Andres Jäschke. Identification and in vitro characterization of UDP-GlcNAc-RNA cap-modifying and decapping enzymes. *Nucleic Acids Research*, 52(10):5438–5450, June 2024. ISSN 0305-1048. doi: 10.1093/nar/gkae353.

Kira J. Weissman and Peter F. Leadlay. Combinatorial biosynthesis of reduced polyketides. *Nature Reviews Microbiology*, 3(12):925–936, December 2005. ISSN 1740-1534. doi: 10.1038/nrmicro1287.

Kehan Wu, Yingce Xia, Pan Deng, Renhe Liu, Yuan Zhang, Han Guo, Yumeng Cui, Qizhi Pei, Lijun Wu, Shufang Xie, Si Chen, Xi Lu, Song Hu, Jinzhi Wu, Chi-Kin Chan, Shawn Chen, Liangliang Zhou, Nenghai Yu, Enhong Chen, Haiguang Liu, Jinjiang Guo, Tao Qin, and Tie-Yan Liu. TamGen: Drug design with target-aware molecule generation through a chemical language model. *Nature Communications*, 15(1):9360, October 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-53632-4.

Kang Zhang, Xin Yang, Yifei Wang, Yunfang Yu, Niu Huang, Gen Li, Xiaokun Li, Joseph C. Wu, and Shengyong Yang. Artificial intelligence in drug development. *Nature Medicine*, 31(1):45–59, January 2025. ISSN 1546-170X. doi: 10.1038/s41591-024-03434-4.

Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, Furui Liu, Guangyong Chen, Chang-Yu Hsieh, and Tingjun Hou. ResGen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, September 2023. ISSN 2522-5839. doi: 10.1038/s42256-023-00712-7.

A Detailed Methods

A.1 Structural Clustering Analysis

We installed the Foldseek (commit `427df8a`) Conda environment to facilitate the computational workflow. To identify structurally homologous protein clusters, we used these `easy-cluster` settings: `--lddt-threshold 0.5` to accept alignments with a local distance difference test (LDDT) score above the specified threshold, and `-e 0.1` where high values indicate more distant structural matches. We selected these values as a balance between sensitivity and specificity (Mifsud et al., 2024; van Kempen et al., 2024). A lower threshold increases cross-species alignments, leading to more clusters that contain human homologs, whereas a higher threshold results in fewer clusters due to increased stringency in alignment scoring.

Proteome Data Acquisition Proteomes were obtained from the AlphaFold Database Proteome Downloads. The following species were selected and retrieved on the respective dates:

- *Escherichia coli*, **reference organism** in our study (November 5, 2024)
- *Enterococcus faecium* (November 6, 2024)
- *Staphylococcus aureus* (November 6, 2024)
- *Klebsiella pneumoniae* (November 6, 2024)
- *Pseudomonas aeruginosa* (November 6, 2024)
- *Helicobacter pylori* (November 7, 2024)
- *Mycobacterium tuberculosis* (November 8, 2024)
- *Homo sapiens* (November 11, 2024)

Data Processing and Essential Gene Selection For essential gene identification, clusters (stored in `*_cluster.tsv`) containing one or more essential genes were selected, and those including any human proteins were excluded. The final dataset consisted of clusters that contained essential bacterial genes with structural homologs in ESKAPE pathogens but without direct human analogs, enabling their potential as antimicrobial targets.

A.2 Structure-based Drug Design

We selected binding site residues of CdsA (modelled using AlphaFold 3 and validated with known substrate-binder structures), MurC (annotation in UniProt), and CohE (dimerization surface) as model input. If coordinates were required, they were determined via center of mass in PyMOL of the relevant residues.

A.2.1 Modifications

ResGen We adapted the ResGen code allowing to set a center without ligand, without changing any implementation details or affecting training weights.

Table A1: All SBDD models used in this work.

Name	Type	GitHub Commit
DiffSBDD	Diffusion	`de4f605`
ResGen	Hierarchical autoregression	`2db3def`
DeepBlock	Building blocks gen., mol. reconstruction	`444a623`
TamGen	Language model	`9459b3f`
TargetDiff	3D equivariant diffusion	`142f1eb`
Pocket2Mol	Equivariant GNN	`836a0c4`

TargetDiff To define the center of the binding pocket, a pseudoatom was created in PyMOL at the desired spatial coordinates. This pseudo-atom was saved as a PDB file and subsequently converted to SDF format using OpenBabel. An index file was then generated. This file contained metadata required for downstream processing, including references to the protein structure (PDB), the pseudo-ligand (SDF), and a placeholder RMSD value. Finally, pocket extraction was performed using a Python script. This script processed each target protein directory, using the provided center and a fixed radius to generate the final pocket representations.

TamGen To prepare the input structures, each target protein file in PDB format was first converted to CIF format. This conversion was performed using an online tool such as the PDBx/mmCIF converter available at `https://mmcif.pdbj.org/converter/index.php?l=en`.

Next, a dataset file in CSV format was created to define the targets. This file included the protein identifier and the three-dimensional coordinates (x, y, z) corresponding to the center of the binding pocket. All target proteins were included in this dataset file.

A custom script was then used to generate the dataset directory structure based on the contents of the CSV file and the corresponding CIF protein files. This process organized the input files into a consistent format expected by downstream processing tools.

Finally, ligand generation was performed for each protein in the dataset using a beam search-based inference script. This procedure was repeated across multiple random seeds to ensure diversity in the generated ligands. The resulting ligand candidates were stored in designated output directories for further analysis.

A.3 Curation

The `SMILESCleaner` class was applied and performs the following cleaning steps:

1. Tries to parse SMILES. (If failed, marks molecules as invalid.)
2. De-duplicates. (If failed, marks molecules as invalid.)
3. Cleans SMILES using the ChEMBL standardizer and returns the parent MolBlock. (If failed, marks molecules as invalid.)
4. Removes all unstable rings. Those are defined as unstable if they occur fewer than 50 times in ChEMBL (Walters, 2024). (If failed, marks molecules as invalid.)
5. Applies REOS filters ("Dundee" set) (Walters et al., 1999). (We modified the code to not return invalid molecules, but the number of REOS alerts.)

A.4 Search for Commercially Available Compounds

We selected six chemical suppliers (Table A2). We identified the most similar using SpaceMACS (Schmidt et al., 2022). The MCS similarity metric favours results with a smaller MCS but a comparable size to the query over those with a larger MCS but a completely different size. This is why we chose MCS similarity instead of size. To identify structurally similar compounds, a chemical space search was performed using the SpaceMACS tool. For each target, the SMILES-derived SDF input file was compared against a molecular space library (`.space` files). The search parameters were configured to return a maximum of five results per query, with a minimum similarity threshold

of 0.5. The search was executed using 64 computational threads to optimize performance, and the verbosity level was set to 4 for detailed logging. Each output was saved as a CSV file named according to the corresponding gene or target.

Table A2: Chemical libraries ("Spaces") and their sizes. Total is not the total number of unique molecules because overlap is not disclosed.

Space	**Molecules**	**Provider**
Freedom Space	1.4×10^{11}	Chemspace
REAL Space	7.0×10^{10}	Enamine
AMBrosia	1.1×10^{11}	Ambinter
eXplore	5.0×10^{12}	eMolecules
GalaXi	1.2×10^{10}	WuXi LabNetwork
CHEMriya	1.2×10^{10}	OTAVA
Total	5.3×10^{12}	

A.5 Molecular Dynamics and Free-Energy Estimates

Protein–ligand complex structures were subjected to explicit-solvent MD simulations followed by end-point binding-energy estimation. Ligand hydrogen atoms were added from the input SMILES using RDKit, after which ligand parameters were assigned with GAFF2 via AmberTools (Case et al., 2023). The protein was parametrized using the ff14SB force field (Maier et al., 2015). Each complex was solvated in an OPCBOX water box with a 10 Å buffer, neutralized, and ionized to 150 mM NaCl. Energy minimization was carried out for 1500 to 5000 steps (system-dependent), followed by a 5 ns heating simulation under constant number of particles, pressure, and temperature (NPT) conditions. The system was gradually heated to 300 K using a Berendsen thermostat (coupling constant 0.1 ps), with initial positional restraints (6.0 kcal/mol/Å^2) applied to non-hydrogen solute atoms and progressively released. Isotropic pressure coupling was maintained using a Berendsen barostat with a relaxation time of 5.0 ps (Berendsen et al., 1984). Production simulations were subsequently carried out for 10 ps under similar NPT conditions, except using a Langevin thermostat (collision frequency 0.5 ps^{-1}) (Loncharich et al., 1992) to maintain 300 K, providing a more realistic sampling of the canonical ensemble. Bonds involving hydrogen atoms were constrained using SHAKE, enabling a 2 fs integration timestep (Ryckaert et al., 1977). Long-range electrostatic interactions were treated using the particle mesh Ewald method (Darden et al., 1993), and a 9.0 Å cutoff was applied to nonbonded interactions. Coordinates, restart files, and energies were written every 10 ps. Binding free energies were estimated from the last 100 frames of each production trajectory using `MMPBSA.py` (Miller et al., 2012), reporting MM/GBSA (Genheden & Ryde, 2015). MM/GBSA employed the Onufriev–Bashford–Case generalized Born model with the modified GB-OBC II radii scaling. Ligand stability within the binding site was quantified by ligand RMSD after alignment of each trajectory frame on the protein backbone.

A.6 Biochemical Experiments

A.6.1 Protein Purification

Recombinant MurC was expressed in *E. coli* BL21(DE3) as a SUMO-fusion construct using a pET28a-SUMO-MurC vector (custom synthesis, *Twist Biosciences*). Cell pellets were thawed on ice and resuspended in lysis buffer (25 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 5 mM DTT, 10 % glycerol, 1 mM PMSF, complete EDTA-free protease inhibitor mini (*Roche*), pH = 8.0). DNase I was added to a final concentration of 0.1 mg mL^{-1}, and cells were lysed by sonication (*Bandelin* SONOPULS equipped with a UW2070 converter and an MS73 lance, 2× [7 min, 30 % intensity, pulse 5; 3 min, 80 % intensity, pulse 5]). The lysate was centrifuged (38 720 g, 30 min, 4 °C), and the supernatant was filtered (0.45 µm GD/X PVDF, *Whatman*). Protein purification was performed using an ÄKTA Purifier 10 (*Cytiva*). The filtrate was loaded onto a 5 mL HisTrap HP column (*Cytiva*), pre-equilibrated in buffer A (25 mM Tris-HCl, 300 mM NaCl, 20 mM imidazole, 5 mM DTT, 10 % glycerol, pH = 8.0). The column was washed with buffer A and buffer B (25 mM Tris-

HCl, 500 mM NaCl, 40 mM imidazole, 5 mM DTT, 10 % glycerol, pH = 8.0). Proteins were eluted using elution buffer (25 mM Tris-HCl, 300 mM NaCl, 300 mM imidazole, 5 mM DTT, 10 % glycerol, pH = 8.0). Fractions were collected and immediately diluted by fractionating into wells pre-filled with imidazole-free buffer A. Suitable fractions were pooled and desalted into desalting buffer (25 mM Tris-HCl, 150 mM NaCl, 20 mM imidazole, 5 mM DTT, 10 % glycerol, pH = 8.0) using 5 mL HiTrap desalting columns (*Cytiva*). The His-SUMO tag was removed by incubation with His-SENP2 at 4 °C overnight. Untagged protein was separated from cleaved tags, uncleaved protein, and His-SENP2 by an additional IMAC purification using two 5 mL HisTrap HP columns pre-equilibrated in desalting buffer, and the flow-through was collected. The flow-through was concentrated using a Vivaspin® centrifugal filter (30 kDa MWCO, *Sartorius*). Size-exclusion chromatography was performed using a HiLoad 16/60 Superdex 200 pg column (*Cytiva*) equilibrated in size-exclusion buffer (20 mM Tris-HCl, 150 mM NaCl, 2.5 mM DTT, 10 mM $MgCl_2$, 10 % glycerol, 0.5 mM ATP, pH = 7.5). Fractions were pooled, and glycerol was added to a final concentration of 20 %. Pooled fractions were concentrated using a Vivaspin® centrifugal filter (30 kDa MWCO, *Sartorius*), aliquoted, flash-frozen in liquid nitrogen, and stored at −80 °C. Protein concentration was determined using a NanoDrop One spectrophotometer and an extinction coefficient of 21 890 M^{-1} cm^{-1} at 280 nm (*ExPASy* ProtParam, https://web.expasy.org/protparam/) (Walker, 2005).

A.6.2 Synthesis of UDP-*N*-acetylmuramic acid (UNAM)

UDP-*N*-acetylmuramic acid was synthesized using recombinant MurA and MurB as described by Weber et al. (2024).

A.6.3 Malachite Green Assay

MurC activity assays were performed based on a previously described procedure with modifications (Hameed P et al., 2014). 1 µL compound stock (100× stock in DMSO) or DMSO was placed in a transparent, flat-bottom 96-well microplate (*Brand*). 89 µL of enzyme mix (tag-free MurC prediluted in assay buffer (25 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM ammonium sulfate, 1.25 mM DTT, 0.002 % Brij-35, pH = 8.0)) or buffer (negative control) were added. In the case of AMP-PCP, 5 µL of a 20× stock solution (dissolved in double-distilled water) was added, followed by 85 µL of a modified enzyme mix supplemented with DMSO to equalize enzyme and DMSO concentrations. The plate was shaken for 10 s and incubated for 15 min at 25 °C. The reaction was initiated by adding 10 µL of substrate mixture containing L-alanine, UNAM, and ATP in assay buffer. The final reaction volume was 100 µL. Final concentrations were 35 nM MurC, 40 µM L-alanine, 40 µM UNAM, and 100 µM ATP. The final DMSO concentration was 1 %. Reactions were carried out for 50 min at 25 °C. Malachite green reagent was freshly prepared by mixing reagent A (0.12 % (w/v) malachite green carbinol hydrochloride in 6 M H_2SO_4), reagent B (7.5 % (w/v) ammonium molybdate tetrahydrate in water), reagent C (11 % (w/v) Tween-20 in water), and water in a ratio of 10:2.5:1:40.5, and filtered through a 0.45 µm GD/X PVDF syringe filter (*Whatman*). 100 µL of malachite green reagent were added, followed by shaking for 10 s and incubation for 30 min at 25 °C. The plate was shaken again, and absorbance at 630 nm was measured using an Infinite F200 Pro plate reader (*Tecan*). Data were normalized to the mean of DMSO-containing samples (defined as 100 % normalized activity, three technical replicates per assay) and DMSO-containing samples lacking MurC (defined as 0 % normalized activity, three technical replicates per assay). Data for the designed compounds were recorded in three replicates originating from three independent experiments, and the normalized data points were subsequently combined. For AMP-PCP, dose-dependent MurC inhibition was exemplified in three technical replicates originating from a single experiment. Data visualization and IC_{50} determination for AMP-PCP were performed using *GraphPad Prism*.

A.6.4 Thermal Shift Assay

1 µL compound stock (50× stock in DMSO) or DMSO was dispensed into a white 96-well PCR plate (*Thermo Fisher Scientific*, #1433529). In the case of AMP-PCP, 1 µL of a 50× stock solution (dissolved in double-distilled water) was dispensed, and a modified enzyme mix supplemented with DMSO was added to equalize DMSO concentrations. 49 µL of enzyme mix containing tag-free MurC and SYPRO Orange (*Thermo Fisher Scientific*) prediluted in MurC assay buffer (25 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM ammonium sulfate, 1.25 mM DTT, 0.002 % Brij-35, pH = 8.0) were added. Final concentrations were 2 µM MurC, 2 % DMSO, and 5× SYPRO Orange. Outer wells

were filled with 50 µL assay buffer to minimize edge effects. The plate was sealed with a clear, adhesive film (*VWR*, #H35100000003100E), and fluorescence was recorded using a CFX96 real-time PCR system (*Bio-Rad*) with a temperature gradient from 4 °C to 89 °C in 0.3 °C increments (5 s per step). Data analysis was performed using the CFX Manager software and *GraphPad Prism*. Relative melting point shifts compared to DMSO-treated MurC (four replicates per assay) were calculated. Data for designed compounds were recorded in four replicates originating from two independent experiments. Further, data for AMP-PCP were recorded in 14 replicates originating from four independent experiments.

B Further Results

B.1 Target Identification

We clustered the predicted proteomes of seven pathogenic bacterial strains of high medical relevance provided on the AlphaFold database (Appendix A.1) using Foldseek, a software for fast structural clustering of large protein structure databases. The selection contained *E. coli, E. faecum, S. aureus, K. pneumoniae, P. aeruginosa, H. pylori,* and *M. tuberculosis*. The resulting clusters were sorted by the number of essential genes, omitting any cluster containing human analogs. Thereby, we identified groups containing genes essential in multiple pathogenic strains, ensuring pan-strain effects, while minimizing toxicity via off-target effects by excluding proteins with homologs in human cells. The most promising clusters are shown in Table B1. Among our identified targets, multiple had recently been published as novel targets for antibiotic development, such as AccD, FtsW and LolE, with the latter resulting in the novel antibiotic "Lolamicin" with a pan-strain Gram-negative activity. These results highlight the potential of our clustering approach, identifying pan-strain antibiotic targets for antibiotics with novel modes of action. Therefore, the clustering approach is a suitable first step of the drug discovery pipeline as it introduces novel targets for application-specific ligand design.

An important factor during target selection for such a pipeline is the structural similarity of our protein targets to the training set of the subsequently applied generative models. Presumably, the closer the targets resemble the training set structures, the better the prediction performance will be. All 6 models use the same CrossDocked2020 dataset as their training set, applying the same sampling and splitting methods. This consistency makes comparisons much easier. While most of these models show convincing case studies in their respective publications, the protein targets closely resemble structures present in the training sets, as seen in Figure B1.

B.2 Molecule Design Models

Recent advances in structure-based drug design have leveraged diverse machine learning (ML) frameworks, each offering distinct strengths and methodologies for molecular design. **DeepBlock** employs a DL approach that decomposes molecules into synthetically accessible and chemically reactive building blocks, enabling targeted ligand synthesis tailored to protein sequences without requiring explicit 3D structural input (Li et al., 2024). Its dual-stage framework integrates a conditional variational autoencoder with a graph-based reconstruction algorithm, uniquely combining biochemical rationale and generative precision.

In contrast, **DiffSBDD** utilizes a 3D SE(3)-equivariant diffusion model, directly generating ligand structures conditioned on protein pocket geometries (Schneuing et al., 2024). It employs symmetry-aware GNNs, ensuring chemically valid and geometrically consistent outputs, and provides versatile generative capabilities adaptable to various design scenarios without retraining.

Similarly, **Pocket2Mol** is an E(3)-equivariant generative model that leverages vector-based neural representations and geometric vector perceptrons to autoregressively sample chemically realistic molecules from pocket geometries (Peng et al., 2022). This method uniquely balances structural accuracy, generative efficiency, and chemical validity, effectively integrating spatial and chemical bonding constraints.

Further enhancing structural fidelity, **ResGen** integrates parallel multiscale modeling within a dual-level autoregressive architecture, capturing both local atomic details and global protein–ligand interactions (Zhang et al., 2023). Its SE(3)-equivariant framework ensures physical plausibility

Table B1: **Essential gene clusters, i.e., potential antibacterial targets.** We used Foldseek clustering to identify novel targets. All clusters have no close human analogs. Based on the scientific literature, we investigated whether clusters are targeted and act as enzymes. Gene names refer to the organism of the representatives. The cluster size n and the number of essential genes n_e within the cluster are shown. Our target selection is highlighted.

Representative	n	$n_e \downarrow$	Protein name	Gene name	Targeted	Catalytic
Q2FXJ0	28	13	UDP-N-acetylmuramate–L-alanine ligase	murC	True	True
Q2G167	206	11	Putative hemin import ATP-binding protein HrtA	hrtA	False	False
P0AFI2	11	6	DNA topoisomerase 4 subunit A	parC	True	True
O25604	12	6	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	accD	True	True
Q2FYS5	11	6	DNA topoisomerase 4 subunit B	parE	True	True
P75974	12	5	Prophage repressor CohE	cohE	False	False
Q9HW01	33	5	UDP-N-acetylglucosamine–N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	murG	True	True
Q2FXY3	13	5	Probable nicotinate-nucleotide adenylyltransferase	nadD	True	True
O33877	11	5	3-hydroxydecanoyl-[acyl-carrier-protein] dehydratase	fabA	True	True
P56042	6	4	Large ribosomal subunit protein bL17	rplQ	—	False
A0A133CSM8	13	4	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	murA	True	False
P75958	21	4	Lipoprotein-releasing system transmembrane protein LolE	lolE	True	False
A0A0H3GRH8	7	4	Outer-membrane lipoprotein LolB	lolB	True	False
O25719	6	4	Riboflavin biosynthesis protein	—	True	True
Q842S4	8	4	Peptide deformylase	def	True	True
A0A1A7SPR9	19	4	HAD family hydrolase	—	False	False
P66572	6	4	Small ribosomal subunit protein uS5	rpsE	—	False
Q2G2C2	11	4	Probable peptidoglycan glycosyltransferase FtsW	ftsW	True	True
Q9HWE4	12	4	Small ribosomal subunit protein uS17	rpsQ	—	False
A0A0H3H383	8	4	7,8-dihydroneopterin aldolase	folB	True	True
O26088	8	4	Organic solvent tolerance-like N-terminal domain-containing protein	—	False	False
A0A133CQC3	7	4	CDP-diacylglycerol–glycerol-3-phosphate 3-phosphatidyltransferase	pgsA	True	True
Q2FW07	6	4	Large ribosomal subunit protein uL4	rplD	—	False
Q9HXH5	8	4	LPS export ABC transporter permease LptG	lptG	True	False
P48940	6	4	Small ribosomal subunit protein uS7	rpsG	—	False
A0A132ZIN5	10	3	—	-	—	-
O25916	10	3	Replicative DNA helicase DnaB	dnaB	True	True
Q2G268	5	3	Coenzyme A biosynthesis bifunctional protein CoaBC	coaBC	True	True
Q9HUL6	6	3	tRNA threonylcarbamoyladenosine biosynthesis protein TsaE	tsaE	False	False
Q9HXJ8	6	3	GTPase Der	der / yhbZ	True	False
A0A132P515	5	3	Large ribosomal subunit protein bL20	rplT	—	False
Q2FYR1	5	3	Aminoacyltransferase FemB	femB	True	True
A0A132Z939	6	3	Chromosomal replication initiator protein DnaA	dnaA	True	False
P0AFG0	7	3	Transcription termination/antitermination protein NusG	nusG	True	False
Q9F1K0	8	3	DNA polymerase III subunit alpha	dnaE	True	True
A0A0H3GZA8	6	3	Large ribosomal subunit protein uL22	rplV	—	False
P76091	8	3	Uncharacterized protein YnbB	ynbB	False	True
P56038	6	3	Large ribosomal subunit protein uL13	rplM	—	False
P56031	5	3	Large ribosomal subunit protein uL3	rplC	—	False
P56036	6	3	Large ribosomal subunit protein uL10	rplJ	—	False
P15042	8	3	DNA ligase	ligA	True	True
P06710	5	3	DNA polymerase III subunit tau	dnaX	False	True
O25688	8	3	Ribosome-binding factor A	rbfA	—	False
P56102	9	3	Methionine aminopeptidase	map	True	True
P08373	5	3	UDP-N-acetylenolpyruvoylglucosamine reductase	murB	True	True
P52097	5	3	tRNA(Ile)-lysidine synthase	tilS	True	True
A0A1X4JGB4	5	3	tRNA (Adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB	tsaB	False	False
Q9HWE2	6	3	Large ribosomal subunit protein uL16	rplP	—	False
A0A133CHI6	11	3	Acyl carrier protein	acpP	—	False
A0A133CVC2	5	3	Translation initiation factor IF-3	infC	True	False

while significantly accelerating the generation of drug-like molecules specifically tailored to protein pockets.

In a different methodological direction, **TamGen** applies a generative pre-trained transformer (GPT)-like chemical language model combined with transformer-based protein encoders to efficiently generate SMILES-based compounds from 3D protein pocket data (Wu et al., 2024). Its architecture supports both *de novo* molecule generation and iterative refinement without relying on explicit 3D ligand structures during training, enhancing chemical validity, synthetic accessibility, and generation speed.

Lastly, **TargetDiff** adopts a non-autoregressive, SE(3)-equivariant diffusion model that jointly models atomic coordinates and types, ensuring efficient and realistic molecule generation conditioned on protein binding sites (Guan et al., 2023). Distinctively, it doubles as an unsupervised feature extractor for binding affinity prediction, integrating molecule generation and evaluation within a unified end-to-end framework.

Overall, these models illustrate diverse yet complementary strategies, varying primarily in their reliance on structural versus sequence data, generative methodologies (autoencoder, autoregressive, diffusion-based), and specific design objectives (efficiency, flexibility, biochemical rationale).

B.3 Phosphate-containing Molecular Designs

After evaluating the initial 44 molecular designs experimentally, we reassessed our post-generation filtering strategy. We noted that the native substrate and reported inhibitors of MurC both contain phosphate groups. This feature is systematically penalized by the REOS/Dundee structural filters and was therefore excluded from our initial selection. Although phosphate groups are generally considered undesirable in drug discovery due to poor permeability and pharmacokinetic liabilities, this criterion can be overly restrictive for enzymes whose catalytic mechanisms engage phosphorylated ligands natively. Furthermore, we observed that the generative models are aware of the biochemical environment of the target. As shown in Figure B5, the phosphor alert is the most common one for MurC. Consequently, we reran the pipeline while explicitly allowing phosphate-containing compounds by disabling the corresponding Dundee alert for this target.

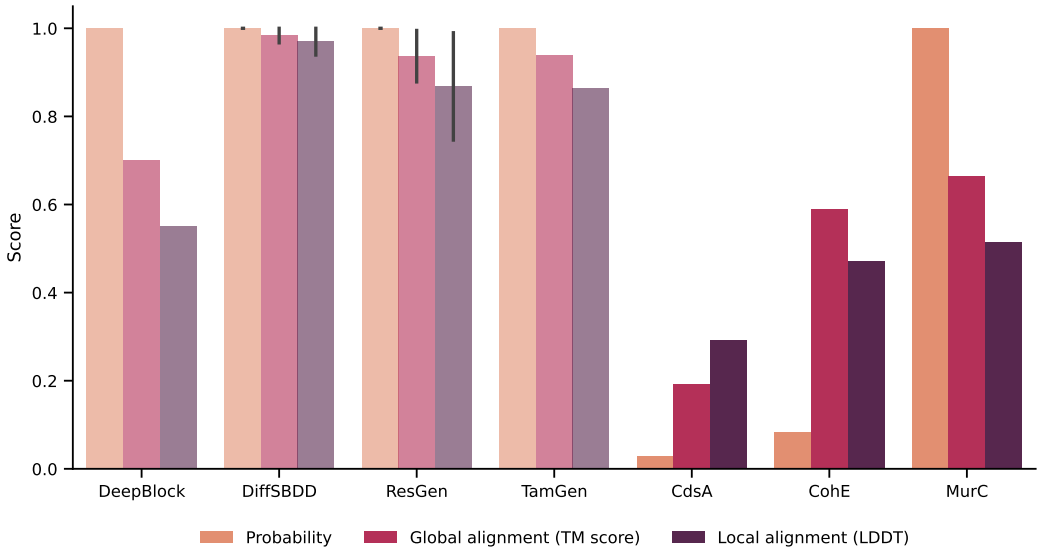

Figure B1: **Foldseek search similarity of target proteins to CrossDocked2020 training set.** We show the top1 result of each Foldseek search for case study proteins from the respective model publications ($\alpha = 0.6$) and our 3 targets ($\alpha = 1$) against the CrossDocked2020 training set. For DiffSBDD and ResGen multiple case study proteins were scored. Pocket2Mol and TargetDiff were omitted as the respective publications show no independent case studies. Three metrics are reported: Probability score (Foldseek's estimated probability of the top1 hit to be homologous to the query structure), the global template modeling score, and LDDT score. The performed case studies per model were: DeepBlock with NCEH1; DifSBDD with BIKE and MPSK1; ResGen with AKT1 and CDK2; TamGen with CLPP2_MYCTU.

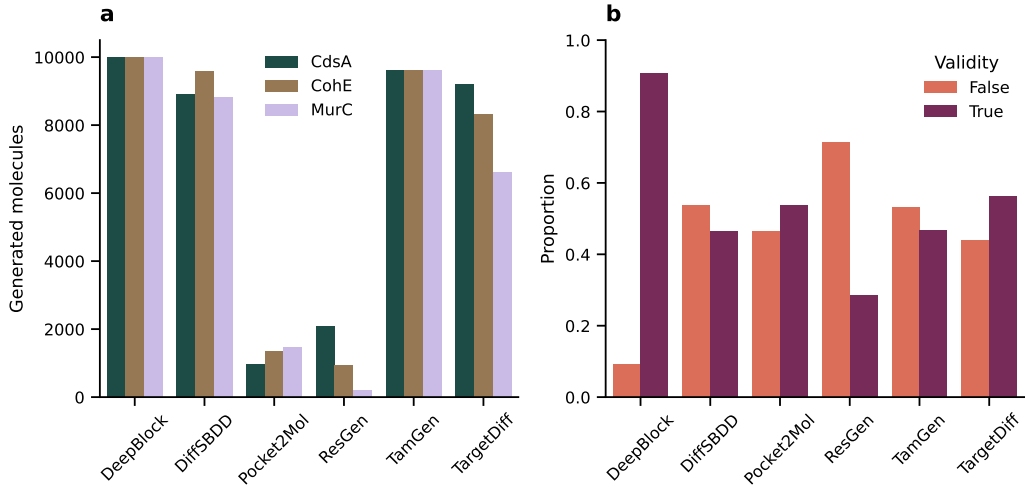

Figure B2: **Overview of generated molecules. a)** We visualize the number of actually generated molecules based on model and target protein. We prompted all models to generate 10 000 molecules per target. **b)** We show the validity of molecules in relation to the SBDD model. The validity is determined by our `SMILESCleaner` class.

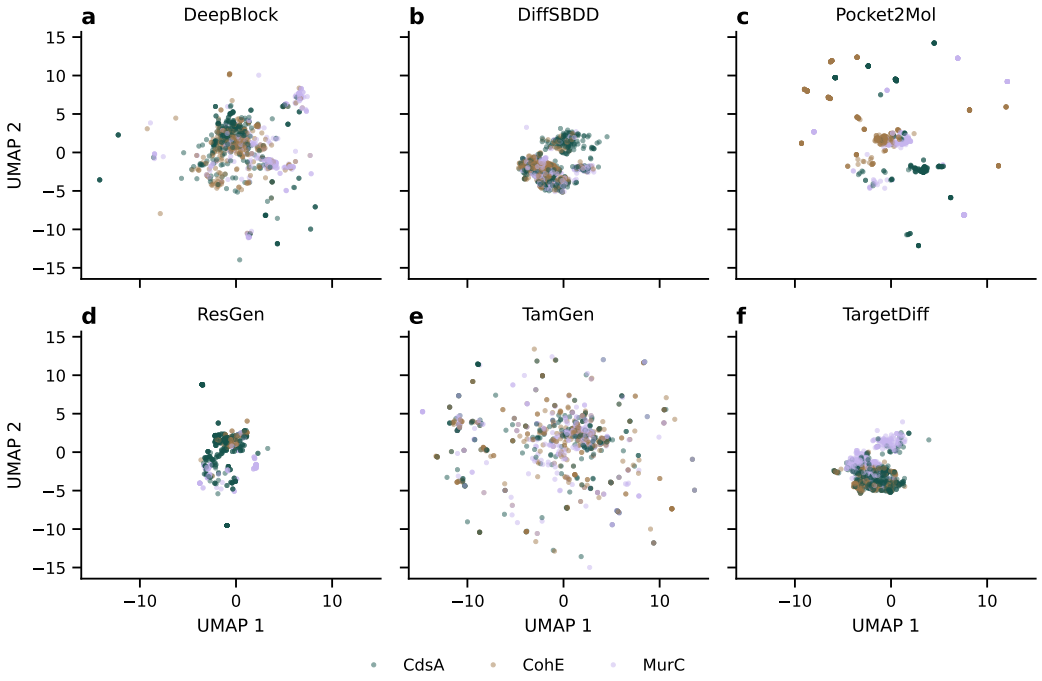

Figure B3: **Model-based chemical space analysis of the generated molecules.** The UMAP space is presented for each model separately (**a–f**), based on ECFPs. Colors represent the different protein targets. We sampled up to 1000 molecules per subplot. Only valid molecules were considered.

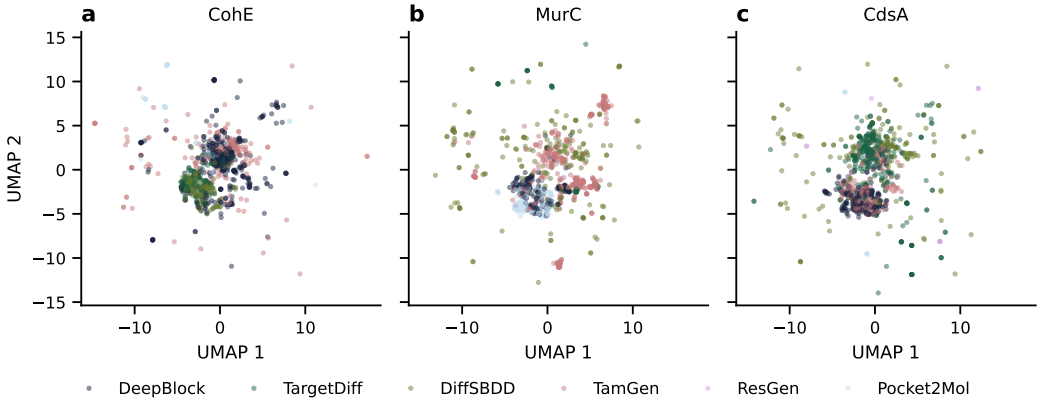

Figure B4: **Target-based chemical space analysis of generated molecules.** The space is presented as a UMAP on a per-target basis (**a–c**) based on ECFPs. Colors represent the different *de novo* model. We sampled up to 1000 molecules per subplot. Only valid molecules were considered.

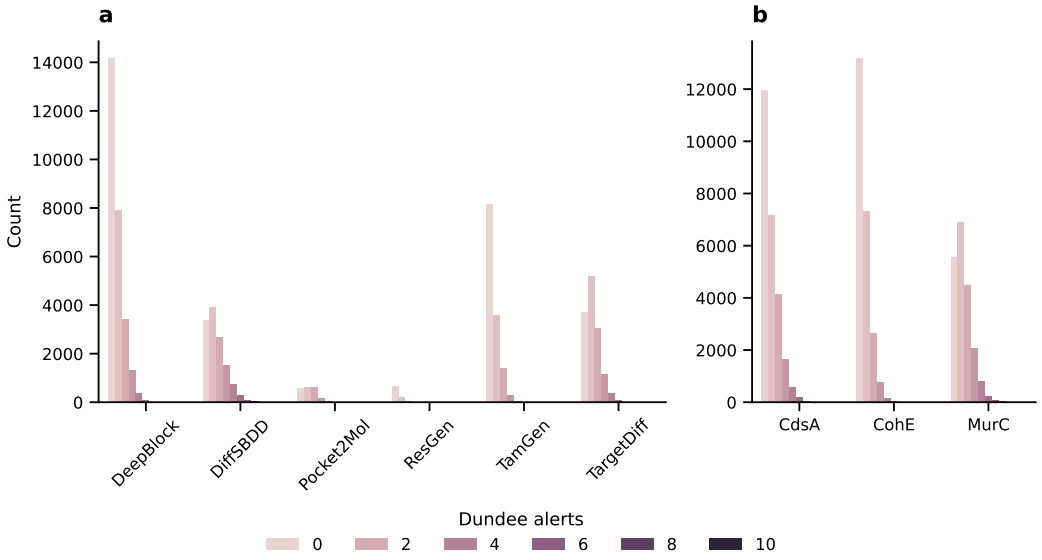

Figure B5: **Structural alerts of generated molecules.** Here, we analyzed the number of Dundee alerts based on model (**a**) as well as target (**b**). Only valid molecules were considered.

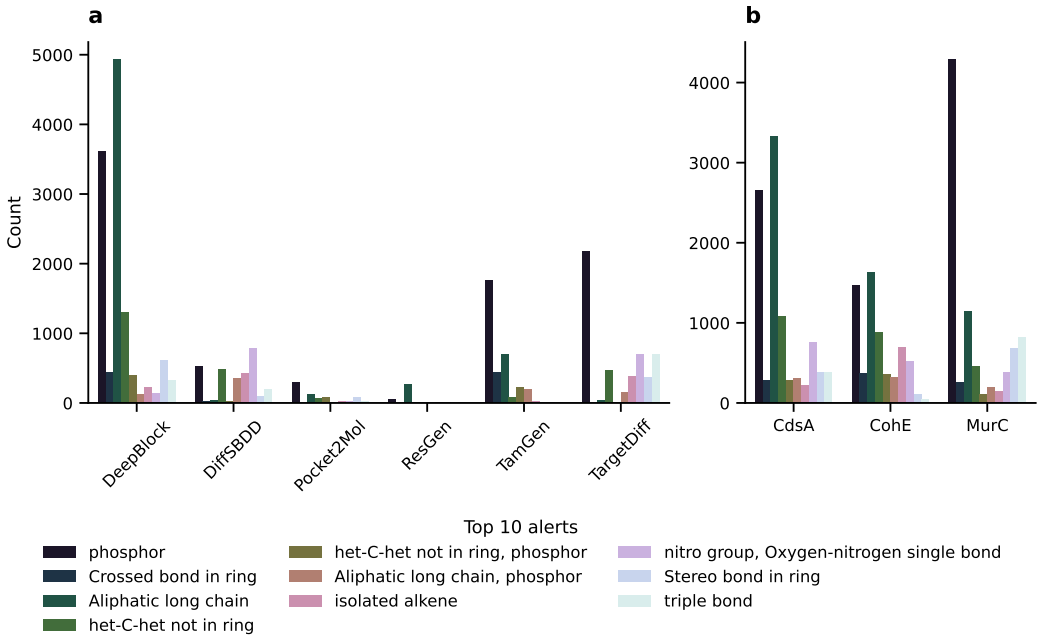

Figure B6: **Top ranking structural alerts of generated molecules.** Here, we investigated which Dundee alerts are the most frequent based on model (**a**) as well as target (**b**). Only valid molecules were considered.

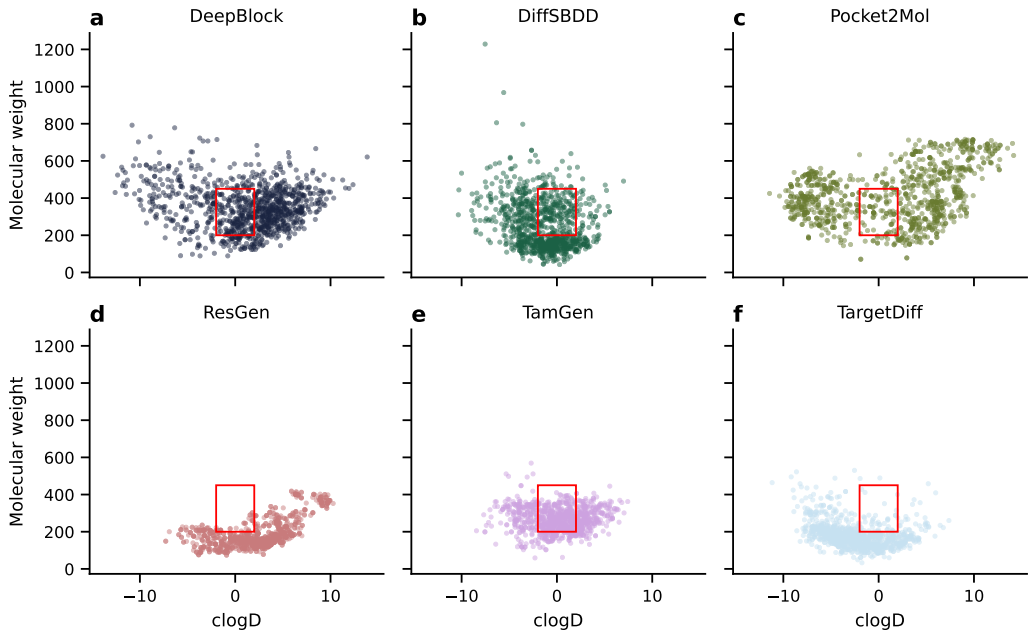

Figure B7: **Property space of generated molecules.** The space is presented separately for each model (**a–f**). Known cytoplasmic antibiotics property space is marked as a box. We visualized molecular weight and calculated logD at pH 7.4. We sampled up to 1000 molecules per subplot. Only valid molecules were considered.

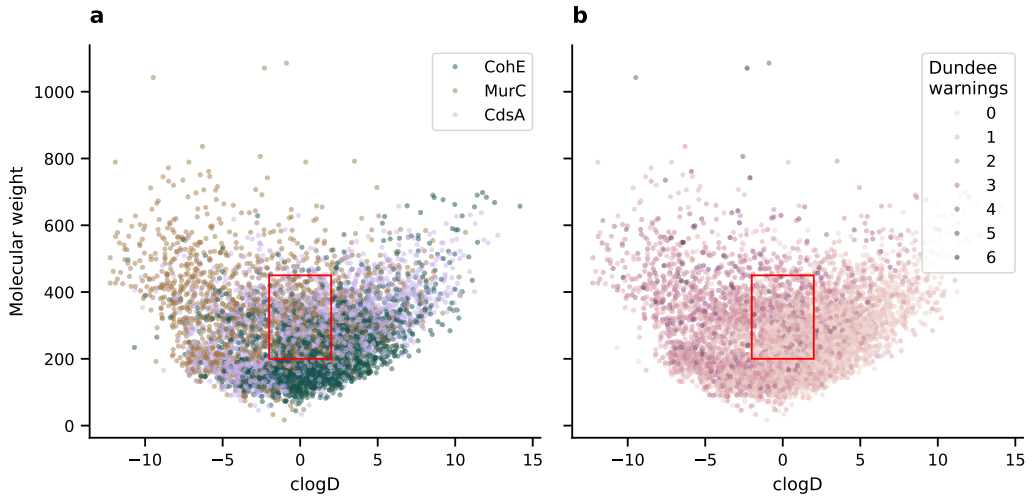

Figure B8: **Property space in relation to target space and structural alerts.** We investigated the distribution in property space dependent on the target (**a**) and the number of Dundee alerts (**b**). Known cytoplasmic antibiotics property space is marked as a box. We visualized molecular weight and calculated logD at pH 7.4. 10 000 molecules were sampled. Only valid molecules were considered.

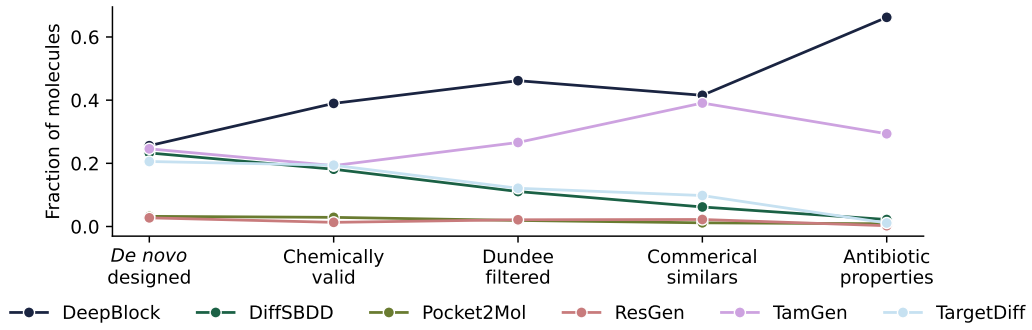

Figure B9: **Processing of *de novo* structures.** We evaluated all models according to the curation and filtering procedure. First, we checked the validity of all generated molecules for the design campaigns against the three targets. Next, we removed any molecules that triggered REOS/Dundee alerts. Then, we assessed how many of the generated scaffolds could be commercially acquired. Finally, we applied an additional filter to select molecules that meet the physicochemical requirements associated with antibiotic activity. Shown here are the relative fractions of remaining molecules per model.

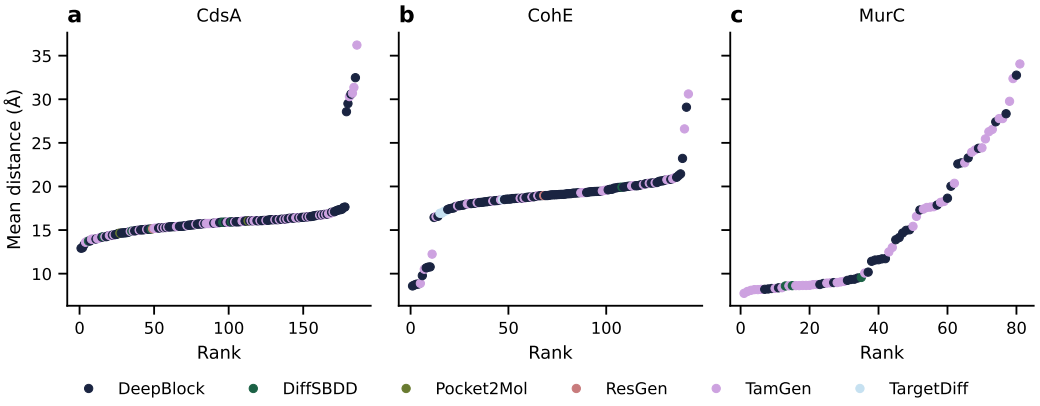

Figure B10: **Distance to pocket based on complex structures.** We predicted complex structures for all selected molecule–target pairs using AlphaFold 3. Then, all molecules were ranked by mean distance. For each protein target (**a–c**), we then calculated the mean distance between the bound ligand and the pocket center, as defined during the SBDD setup. Only ligands within the 90th percentile were kept for further ranking, keeping ligands close to the desired position.

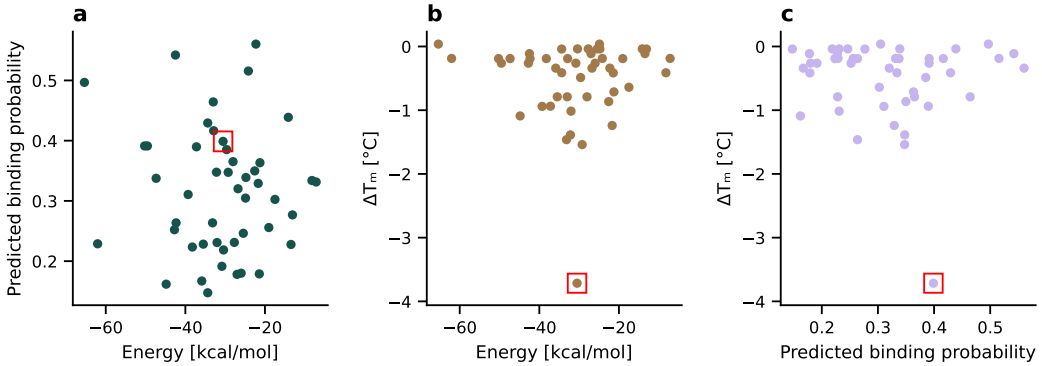

Figure B11: **Correlations of scoring methods and experimental data.** The scoring results between Boltz 2 affinity probability and MM/GBSA scores for the selected 47 compounds showed no correlation (**a**). Neither MM/GBSA (**b**) nor Boltz 2 (**c**) identified the active compound C_{20} out of all 47 candidates. The box indicates C_{20}.

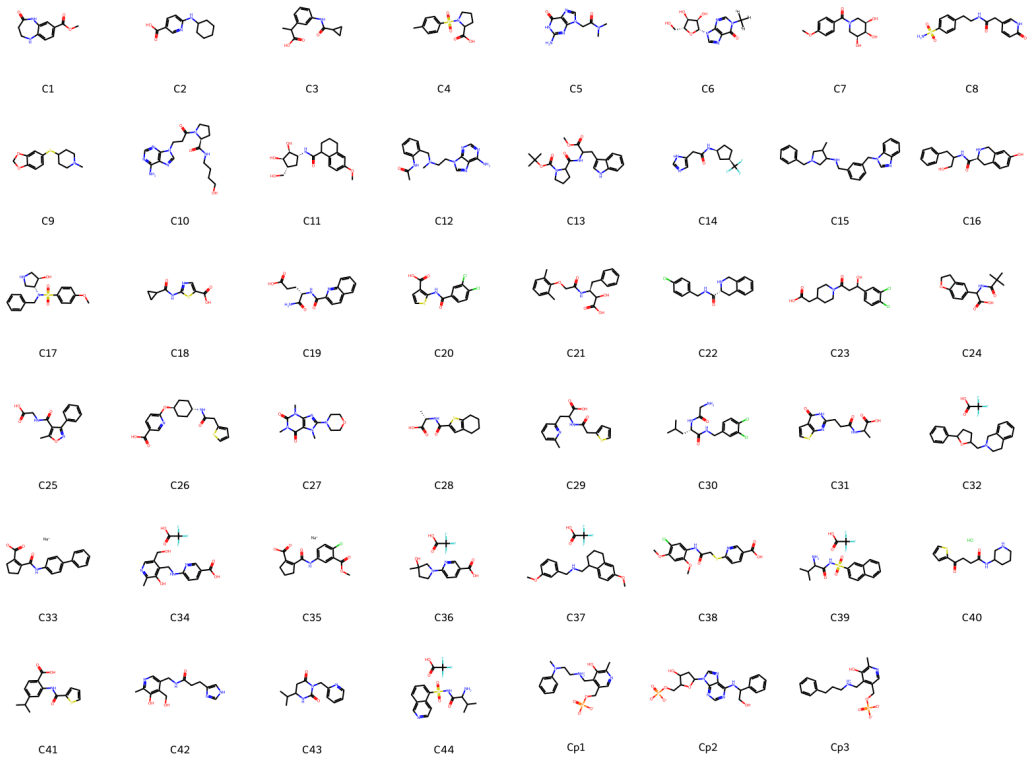

Figure B12: **Structures of 47 experimentally tested compounds.** These 47 compounds were acquired from Enamine and eMolecules. Compounds Cp1–3 contain phosphates and were added despite triggering Dundee alerts due to MurC's affinity to phosphate ligands.

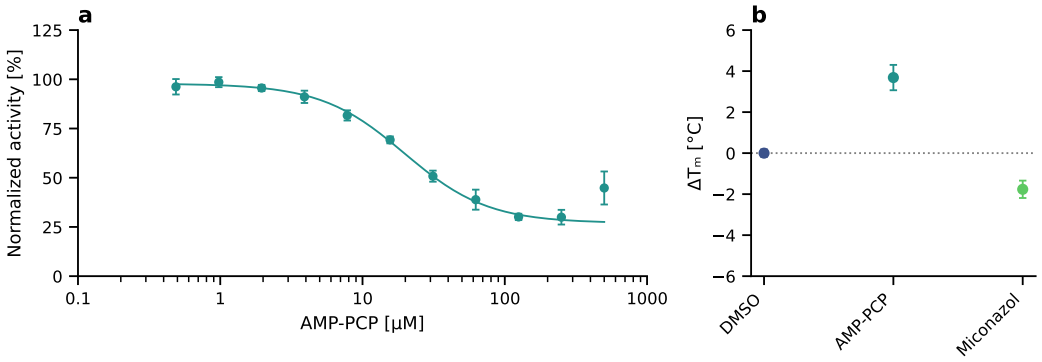

Figure B13: **Control experiments for malachite green assay and thermal shift assay.** The known MurC inhibitor AMP-PCP shows concentration-dependent inhibition of MurC in the malachite green assay (**a**). AMP-PCP also shows strong thermal stabilization of MurC at 200 µM (**b**). The known aggregator Miconazol shows moderate protein destabilization at 200 µM.