SYNLAD: LATENT DIFFUSION FOR GENERATING SYNTHESIZABLE MOLECULES CONDITIONED ON 3D PHARMACOPHORE PROFILES

Anonymous authors
Paper under double-blind review

ABSTRACT

We present SynLaD, a latent diffusion framework for small-molecule generation that unifies 3D design objectives (what to make) with synthetic accessibility (how to make it). Current models typically optimize one objective at the expense of the other, creating a bottleneck for discovering high-scoring and experimentally testable molecules. SynLaD combines reaction-constrained generation with pharmacophore-conditioned 3D design by learning a latent space that decodes to both 3D structures and synthesis pathways. An encoder maps molecules to a latent representation used by two decoder heads: (i) a geometric head that reconstructs atom types and coordinates and (ii) an autoregressive synthesis head that outputs synthetic routes in a serialized, reaction-based notation. A diffusion transformer generates novel latents in the learned space, conditioned on pharmacophore profiles. Across analogue-generation tasks for bioactive ligands, SynLaD outperforms existing baselines in synthesizable and diverse *hit* generation, demonstrating that a single model can produce shape-accurate molecules with feasible synthesis plans.

1 INTRODUCTION

Ligand-based drug design (LBDD) uses known bioactive compounds to guide the design of new molecules with similar three-dimensional (3D) shapes and physicochemical properties (Acharya et al., 2011). Central to LBDD is the pharmacophore, which specifies the spatial arrangement of interaction features—such as hydrogen bond donors or acceptors, charged groups, and aromatic centers—responsible for protein–ligand complementarity.

A common LBDD strategy extracts pharmacophores from known ligands and uses them in virtual screening to identify molecules with similar shapes and features (Goodnow & Gillespie, 2007). However, the rapid expansion of chemical space (Kuan et al., 2023) renders exhaustive screening increasingly computationally prohibitive. Generative modeling has emerged as a promising alternative to brute-force search, allowing for faster inference times and the proposing of molecules outside of pre-defined spaces (Peng et al., 2022; Luo et al., 2021; Schneuing et al., 2024; Guan et al., 2023; Cremer et al., 2024; Schneuing et al., 2025; Cremer et al., 2025; Zhu et al., 2023). While performing well at capturing distributions of desired molecular motifs, these models often produce synthetically inaccessible molecules (Gao & Coley, 2020; Stanley & Segler, 2023).

Synthesis-aware generative methods for designing new molecules have typically either treated synthesizability as an additional property to optimize (Liu et al., 2022; Guo & Schwaller, 2025; Gómez-Bombarelli et al., 2018) or enforced it by generating molecules via synthetic pathways/reactions rather than as unconstrained molecular graphs (Hartenfeller et al., 2012; Vinkers et al., 2003; Horwood & Noutahi, 2020; Bradshaw et al., 2019b; Gao et al., 2022; Swanson et al., 2024; Cretu et al., 2025). However, these approaches are not conditioned on three-dimensional chemical features, which are critical for protein–ligand binding. As a result, bioactive ligand generation is typically framed either as a reward-based problem using binding-affinity proxies (Horwood & Noutahi, 2020; Cretu et al., 2025) or as conditioning on known ligands via chemical space projection (Luo et al., 2024).

We bridge the gap between 3D chemical feature conditioning and synthesizable molecule generation with SynLaD (*Synthesis-aware Latent Diffusion*), a conditional latent diffusion model that

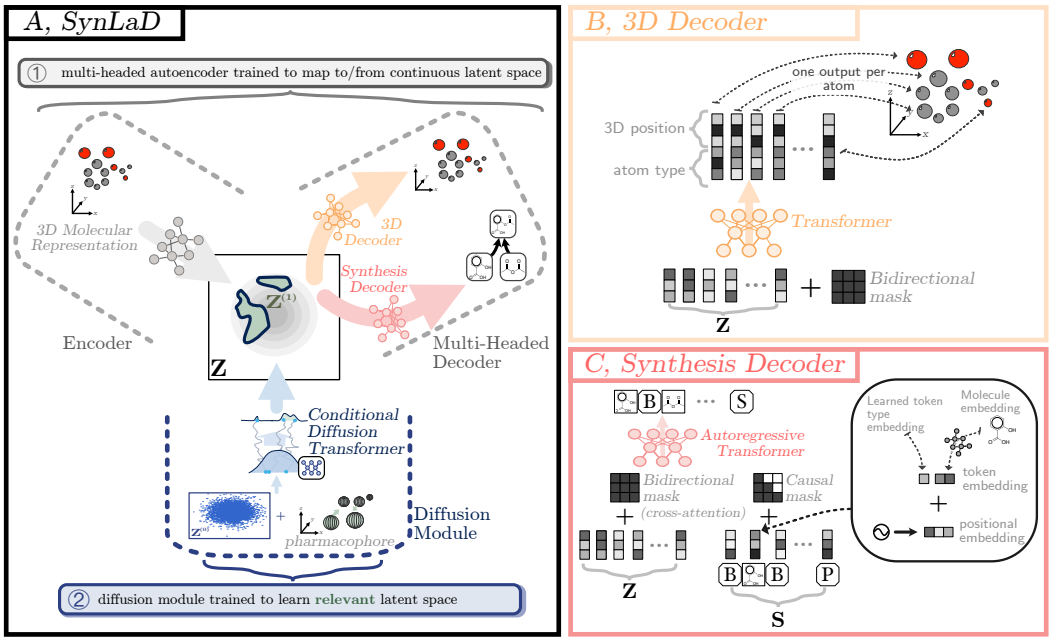

Figure 1: **A, Overview of SynLaD: a multi-head generative model that jointly produces 3D molecular structures and synthesis plans.** SynLaD is trained in two stages: ① an autoencoder learns a continuous latent representation of 3D molecules, with an auxiliary decoder that generates a synthesis plan for each structure. ② a pharmacophore-conditioned diffusion model learns to sample latent representations that decode to molecules containing a desired pharmacophore. **B, 3D Decoder:** a transformer with a bidirectional mask maps latent representations to atom types and 3D coordinates. **C, Synthesis Decoder:** an autoregressive transformer with cross-attention to the latent space decodes latent representations into serialized synthesis plans.

decodes latents to molecules in both 3D and synthesizable space. An overview of SynLaD is shown in Figure 1. At its core, SynLaD employs a molecular variational autoencoder with a dual-decoder architecture: a single encoder maps a 3D molecule to the latent space, from which one decoder reconstructs the 3D atomic coordinates (and corresponding atom types), while a second autoregressive decoder generates a corresponding synthesis pathway. Joint training of these decoders couples spatial fidelity and synthetic viability in the learned latent representation. Conditional generation is then performed in the learned latent space using a Diffusion Transformer (DiT) (Rombach et al., 2022; Joshi et al., 2025). Importantly, operating in latent space enables the generation of rich embeddings that are decoded by the synthesis head. This introduces variability (i.e. different conditioning inputs to the synthesis decoder), which we demonstrate is essential for ensuring sample diversity.

We evaluate SynLaD using standard molecule quality metrics, including pose validity and synthesizability, under both unconditional and pharmacophore-conditioned generation. We further assess its practical utility on bioactive analogue generation for ligands from the Lit-PCBA benchmark (Tran-Nguyen et al., 2020). Across settings, SynLaD matches the molecule-quality profile of its separately-trained counterparts while consistently improving the fraction of samples with plausible synthetic routes, and it outperforms existing baselines in producing synthesizable, diverse hits under pharmacophore-guided sampling.

2 SynLaD–Pharmacophore-conditioned ligand generation

We present SynLaD—a latent diffusion (Rombach et al., 2022) model for *de novo* small molecule generation conditioned on pharmacophore features (see Figure 1 for an overview and Appendix C.2 for a full description of our method). SynLaD is trained in two stages: In stage 1, a variational autoencoder (Kingma & Welling, 2013; Rezende & Mohamed, 2015) encodes molecules into a shared latent space and a two-headed decoder reconstructs both the 3D molecular structure and a synthesis pathway. In stage 2, we train a Diffusion Transformer (Peebles & Xie, 2023) to generate new samples from the latent space, which are decoded into both 3D molecules and synthesis pathways.

The two decoders are trained to reconstruct the *same* molecule under two *different* representations and at inference time the synthesis decoder is used to ensure synthesizability of generated designs. The synthesis decoder uses a serialized representation of chemical reactions encoded using directed acyclic graphs, and a causal transformer to generate sequences using a custom embedding scheme. A reaction prediction oracle is used at test time to infer products from the generated reactant tokens (Schwaller et al., 2019; Bradshaw et al., 2020). We train a conditional flow matching model (Lipman et al., 2023; Song & Ermon, 2019; Ho et al., 2020; Joshi et al., 2025) on the latent space learned by the autoencoder, and embed pharmacophores via a transformer trained jointly with the denoiser backbone. We detail the autoencoder and denoiser design in Appendix C.2.

3 EXPERIMENTS

Dataset. We train SynLaD on a set of reaction pathways extracted from the USPTO dataset (Lowe, 2017). We detail the construction of the dataset in Appendix C.1. Our dataset contains 67,512 synthesis pathways, each with 1 to 6 intermediary reactions. For each final product of the pathways, we compute 10 low-energy conformers using OpenEye's conformer generation software Omega (Hawkins et al., 2010; Friedrich et al., 2017). Our dataset contains synthesis products with 10 to 40 heavy atoms. Metrics definitions and evaluation are described in Appendix C.10.

As described in Section 2, we use a reaction-prediction oracle to infer product information from sets of reactant tokens sampled by our decoder. The oracle is trained separately; implementation details and accuracy results are provided in Appendix C.9, where it achieves a top-1 product prediction accuracy of 84.6%.

3.1 JOINT CONFORMER AND SYNTHESIS LATENT SPACE

We first investigate reconstruction performance of the joint autoencoder. We randomly select 1,000 molecules from the test set, encode them into our latent space, and decode each latent using both decoders. For the 3D decoder, we report match rate and RMSD between ground-truth and predicted coordinates (see Appendix C.10); for the synthesis decoder, we report match rate and 2D fingerprint Tanimoto similarity between the predicted end product and the ground-truth synthesis outcome. We ablate the latent dimension during training and compare inference strategies for the synthesis decoder (see Appendix C.5). Results are summarized in Table 4 for both *sampling* and *beam search* decoding. *Sampling* generates N independent candidate sequences by temperature-scaled stochastic sampling at each step, optionally using *top-k* truncation by setting logits of tokens outside of the top k to $-\infty$. In contrast, *beam search* deterministically expands the b (number of beams) most probable partial sequences. In our experiments, we set $N = 1$ and $k = 10$ for sampling and compare against beam search with $b = 5$. Beam search consistently outperforms top-k sampling and is therefore used for latent-space sampling in subsequent experiments.

3.2 UNCONDITIONAL GENERATION

We next train an unconditional latent DiT denoiser using latents produced by our joint VAE described in Section 3.1; model parameters and training details are provided in Appendix C.7. We sample using $T = 100$ ODE integration steps and decode the resulting latent with both decoders. Results are reported in Table 1, where we compare against (1) a latent diffusion model trained without a synthesis decoder head (equivalent to Joshi et al. (2025)) and (2) an unconditional autoregressive synthesis-decoder-only head. The results show that jointly training the two decoders does not degrade synthesizability or PoseBusters performance, and yields good validity, diversity, and novelty overall. Interestingly, joint training improves the synthesizability of molecules decoded by the 3D decoder, suggesting that the synthesis decoder steers generation toward latent regions more likely to decode to synthesizable molecules. Both SynLaD decoders significantly outperform the synthesis-unconstrained baseline in synthesizability although trained on the same data.

3.3 PHARMACOPHORE-CONDITIONED MOLECULE GENERATION

In-distribution pharmacophore conditioning. We evaluate the conditional generation performance of SynLaD by selecting 50 random *query* molecules from the test set and extracting their pharmacophores. For each query, we generate 100 candidate molecules with SynLaD conditioned on the extracted pharmacophore, and, as a baseline, compare against 5,000 randomly selected molecules from the training set. All candidates–whether sampled from the training set or generated by SynLaD–are evaluated using conformer enumeration with Omega, followed by shape and

Table 1: **Unconditional generation results.** ↑/↓ indicate that higher/lower is better. All metrics (see Appendix C.10) are evaluated over 1,000 samples, except for AiZynthFinder (100 samples).

	Method	Valid.↑	IntDiv.↑	Nov.↑	FCD↓	PB↑	SA↓	AiZynth.↑
separately-trained heads	3D Decoder	88.5	86.0	100	2.91	0.89	2.78	0.59
	Synthesis Decoder	100	88.9	52.7	2.10	-	2.30	0.76
jointly-trained heads	SynLaD (3D outputs)	90.0	86.0	100	2.70	0.89	2.65	0.72
	SynLaD (synthesis outputs)	100	88.2	84.6	1.52	-	2.30	0.87

pharmacophore overlay scoring against the query molecule (see Appendix C.11). For each molecule, we retain the highest scoring conformer and assess whether it is a hit or not (i.e., with a Tanimoto combo score ≥ 1.2 to the query; for metrics definitions see Appendix C.10). We report the median number of hits and unique scaffold hits in Figure 2 and distributions of Tanimoto shape and color scores, as well as synthesizability scores assessed using AiZynthFinder (further metrics in Appendix Table 8). Results show that SynLaD generates samples with significantly higher 3D pharmacophore and shape similarities to the query than the baseline, while the synthesis decoder yields molecules with high synthesizability.

Method	Hits	Unique scaff. hits	Max score	AiZynth.	Synthesizable hits
Dataset baseline	0	0	1.14	0.90	0
SynLaD (3D out)	36.5	13.5	1.92	0.44	18.5
SynLaD (syn out)	29	9.5	1.88	0.80	25.0

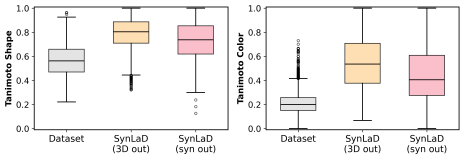

Figure 2: **Pharmacophore conditioned generation.** *Left*: Values for hits, unique scaffold hits, and max score are medians across the 50 query molecules, while AiZynth is a mean. Max score represents the maximum Tanimoto combo score for generated samples. *Right*: Distributions of ROCS shape and color (pharmacophore) similarity scores to query molecules (higher better).

Screening case study. To assess SynLaD in a practical drug-discovery setting, we compare it against brute-force ROCS screening over our dataset, which serves as a proxy for a molecular library and reflects standard but time-consuming ligand-based screening. We randomly select 10 query molecules with more than 15 heavy atoms from our test set and calculate ROCS scores against all other molecules in the dataset (∼67k molecules), excluding the query molecules themselves. In parallel, we perform pharmacophore-conditional generation with SynLaD, generating 1,000 samples per query. For this experiment, we consider only outputs of the synthesis decoder of SynLaD, ensuring that outputs are synthesizable and directly applicable for experimental testing. We report hit counts for both methods in Table 9, and per-query results in Appendix Figure 7. SynLaD achieves higher hit counts than the typical brute-force baseline in ligand-based screening, while using nearly two orders of magnitude fewer samples (1k molecules) than the size of the screened library (∼67k).

Table 2: **Bioactive hit diversification.** We sample 500 samples per query for all methods. We report averaged metrics over all ten queries and AiZynthFinder success rate for a random subset of 100 generated molecules. Amortized methods do not require retraining for each target. Max score refers to the maximum Tanimoto combo score registered.

Method	Validity	Hits (avg.)	Unique scaff. hits (avg.)	Max score	AiZynth.	Num. queries w/ $\geq$ 1 hit	Amortized	Synthesizable hits (avg.)
REINVENT	0.98	21.1	3.7	1.15	**0.75**	3	✗	**16.3**
ShEPhERD	0.55	15.7	13.0	1.23	0.18	6	✓	6.3
SynFormer	**1.0**	13.8	6.7	1.31	0.50	**8**	✓	6.7
SynLaD (3D out)	0.49	**38.1**	**18.8**	**1.45**	0.22	**8**	✓	12.7
SynLaD (syn out)	**1.0**	17.9	6.9	1.30	**0.75**	5	✓	14.6

Bioactive hit diversification. We next investigate the out-of-distribution performance of SynLaD on a bioactive hit diversification task, which aims to find structural analogues of known bioactive compounds. This mirrors real-world drug discovery practices, where diverse candidates are needed to iteratively optimize activity while satisfying synthesizability, developability, and other constraints. We select ten targets from the Lit-PCBA benchmark (Tran-Nguyen et al., 2020) (see Appendix D.4)

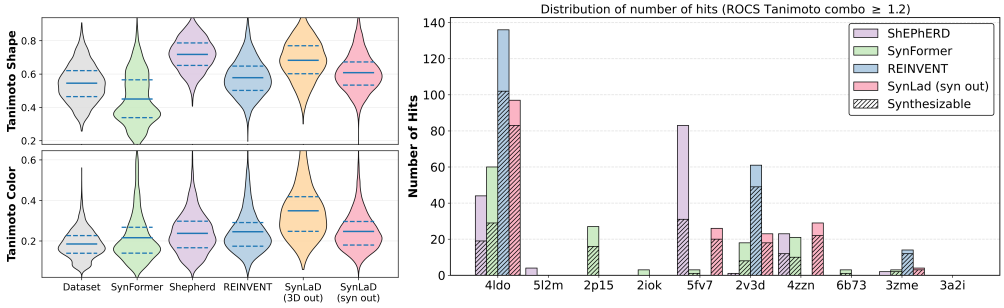

Figure 3: **Bioactive hit diversification.** Structural analogues are generated for 10 Lit-PCBA ligands using SynLaD and baseline methods. We report (1) the distribution of aggregated Tanimoto shape and color similarity scores to the query, and (2) the number of hits per query for each method, highlighting the fraction of synthesizable hits.

and compare against ShEPhERD (Adams et al., 2025), SynFormer (Gao et al., 2022), and REINVENT (Blaschke et al., 2020; Loeffler et al., 2024) (see Appendix D.6).

For each query, we generate 500 molecules by conditioning on the pharmacophore profile of the native ligand in its bound conformation. We apply the same evaluation pipeline as in the previous section and report distributions of aggregated Tanimoto shape and color scores across all ten queries, along with averaged molecule validity, synthesizability, and hit counts, in Figure 3 and Table 2. Note that AiZynthFinder is reported for subsets (100 samples due to compute cost) of all generated molecules, not just hits. SynLaD achieves a strong overall balance, producing a high number of synthesizable hits, succeeding on a large fraction of queries, and maintaining high validity. Compared to the other synthesis-constrained baseline, SynFormer, SynLaD's synthesis decoder produces candidates with consistently higher ROCS shape and color overlap, suggesting that the explicit 3D representation our model learns improves conditional design; it also yields more hits and higher retrosynthesis success. The 3D outputs of SynLaD outperform ShEPhERD's on most queries in terms of pharmacophore overlap, hit count, and synthesizability, while achieving comparable shape overlap. The synthesis-decoder output also outperforms ShEPhERD in average number of synthesizable hits. Although REINVENT achieves the highest average number of synthesizable hits, this result is dominated by a single query with an unusually large hit count; across the benchmark, it generates synthesizable hits for only three queries and shows limited scaffold diversity. We show examples of query-generated molecule pairs with highlighted pharmacophore and shape overlap in Appendix Figures 8 and 9.

While we report results for both decoders for completeness, the key strength of our approach lies in the synthesis decoder, which—despite operating on linear synthesis trajectories—learns to satisfy shape and pharmacophore constraints more effectively than baselines that do not learn 3D information. This highlights the benefit of our strategy: learning latent representations that encode rich 3D conditioning signals and using them to guide synthesis generation. We further support this claim in Appendix D.5, where ablating the 3D branch of SynLaD while keeping the synthesis architecture fixed leads to a substantial performance drop, demonstrating that the 3D component is critical.

4 Conclusion

In this work, we introduce a conditional latent diffusion framework that decodes latent representations into both 3D molecular structures and reaction-based synthesis pathways. We show that jointly training these two decoders is mutually beneficial at sampling time: it increases the synthesizability of molecules produced by the 3D decoder and improves the synthesis decoder's ability to satisfy shape/pharmacophore constraints under conditioning, while also enchancing sample diversity. Overall, our results demonstrate that conditioning an autoregressive synthesis decoder through a 3D-informed latent representation yields high numbers of diverse, synthesizable, pharmacophore-aligned hits. Looking ahead, an exciting direction is to scale our method to larger datasets (e.g., Pistachio, NextMove Software, 2025), which would enable access to more novel transformations and more combinatorially complex regions of chemical space.

REFERENCES

C. Acharya, A. Coop, J. E. Polli, and A. D. Mackerell, Jr. Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach. *Current Computer-Aided Drug Design*, 7(1):10–22, 2011. URL https://doi.org/10.2174/157340911793743547.

Keir Adams, Kento Abeywardane, Jenna Fromer, and Connor W. Coley. ShEPhERD: Diffusing shape, electrostatics, and pharmacophores for bioisosteric drug design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=KSLkFYHlYg.

Guy W. Bemis and Mark A. Murcko. The properties of known drugs. 1. Molecular frameworks. *Journal of Medicinal Chemistry*, 39(15):2887–2893, 1996. URL https://doi.org/10.1021/jm9602928.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003. URL https://dl.acm.org/doi/10.5555/944919.944966.

Hangrui Bi, Hengyi Wang, Chence Shi, Connor Coley, Jian Tang, and Hongyu Guo. Non-autoregressive electron redistribution modeling for reaction prediction. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 904–913. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/bi21a.html.

Thomas Blaschke, Josep Arús-Pous, Hongming Chen, Christian Margreitter, Christian Tyrchan, Ola Engkvist, Kostas Papadopoulos, and Atanas Patronov. REINVENT 2.0: An AI tool for de novo drug design. *Journal of Chemical Information and Modeling*, 60(12):5918–5922, 2020. URL https://doi.org/10.1021/acs.jcim.0c00915.

John Bradshaw, Matt J. Kusner, Brooks Paige, Marwin H. S. Segler, and José Miguel Hernández-Lobato. A generative model for electron paths. In *International Conference on Learning Representations*, 2019a. URL https://openreview.net/forum?id=r1x4BnCqKX.

John Bradshaw, Brooks Paige, Matt J Kusner, Marwin Segler, and José Miguel Hernández-Lobato. A model to search for synthesizable molecules. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019b. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/46d0671dd4117ea366031f87f3aa0093-Paper.pdf.

John Bradshaw, Brooks Paige, Matt J Kusner, Marwin Segler, and José Miguel Hernández-Lobato. Barking up the right tree: An approach to search over molecule synthesis DAGs. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6852–6866. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4cc05b35c2f937c5bd9e7d41d3686fff-Paper.pdf.

John Bradshaw, Anji Zhang, Babak Mahjour, David E Graff, Marwin HS Segler, and Connor W Coley. Challenging reaction prediction models to generalize to novel chemistry. *ACS Central Science*, 11(4):539–549, 2025. URL https://doi.org/10.1021/acscentsci.5c00055.

Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chemical Science*, 15:3130–3139, 2024. URL http://dx.doi.org/10.1039/D3SC04185A.

Julian Cremer, Tuan Le, Frank Noé, Djork-Arné Clevert, and Kristof T. Schütt. PILOT: Equivariant diffusion for pocket-conditioned de novo ligand generation with multi-objective guidance via importance sampling. *Chemical Science*, 15:14954–14967, 2024. URL http://dx.doi.org/10.1039/D4SC03523B.

Julian Cremer, Ross Irwin, Alessandro Tibo, Jon Paul Janet, Simon Olsson, and Djork-Arné Clevert. FLOWR: Flow matching for structure-aware de novo, interaction- and fragment-based ligand generation, 2025. URL https://arxiv.org/abs/2504.10564.

Miruna Cretu, Charles Harris, Ilia Igashov, Arne Schneuing, Marwin Segler, Bruno Correia, Julien Roy, Emmanuel Bengio, and Pietro Lio. SynFlowNet: Design of diverse and novel molecules with synthesis constraints. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=uvHmnahyp1.

Kien Do, Truyen Tran, and Svetha Venkatesh. Graph transformation policy network for chemical reaction prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 750–760, New York, NY, USA, 2019. Association for Computing Machinery. URL https://doi.org/10.1145/3292500.3330958.

Ian Dunn and David Ryan Koes. Mixed continuous and categorical flow matching for 3D de novo molecule generation, 2024. URL https://arxiv.org/abs/2404.19739.

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(8):1–11, 2009. URL https://doi.org/10.1186/1758-2946-1-8.

David Fooshee, Aaron Mood, Eugene Gutman, Mohammadamin Tavakoli, Gregor Urban, Frances Liu, Nancy Huynh, David Van Vranken, and Pierre Baldi. Deep learning for chemical reaction prediction. *Molecular Systems Design & Engineering*, 3(3):442–452, 2018. URL https://doi.org/10.1039/C7ME00107J.

Nils-Ole Friedrich, Christina de Bruyn Kops, Florian Flachsenberg, Kai Sommer, Matthias Rarey, and Johannes Kirchmair. Benchmarking commercial conformer ensemble generators. *Journal of Chemical Information and Modeling*, 57(11):2719–2728, 2017. URL https://doi.org/10.1021/acs.jcim.7b00505.

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. URL https://doi.org/10.1021/acs.jcim.0c00174.

Wenhao Gao, Rocío Mercado, and Connor W. Coley. Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=FRxhHdnxt1.

Wenhao Gao, Shitong Luo, and Connor W. Coley. Generative AI for navigating synthesizable chemical space. *Proceedings of the National Academy of Sciences*, 122(41):e2415665122, 2025. URL https://www.pnas.org/doi/abs/10.1073/pnas.2415665122.

S. Genheden, A. Thakkar, V. Chadimová, J. L. Reymond, O Engkvist, and Esben Bjerrum. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *Journal of Cheminformatics*, 12:70, 2020. URL https://doi.org/10.1186/s13321-020-00472-1.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL http://www.deeplearningbook.org.

R. Goodnow, Jr. and P. Gillespie. 1 Hit and lead identification: Efficient practices for drug discovery. volume 45 of *Progress in Medicinal Chemistry*, pp. 1–61. Elsevier, 2007. URL https://doi.org/10.1016/S0079-6468(06)45501-6.

Sai Krishna Gottipati, Boris Sattarov, Sufeng Niu, Yashaswi Pathak, Haoran Wei, Shengchao Liu, Shengchao Liu, Simon Blackburn, Karam Thomas, Connor Coley, Jian Tang, Sarath Chandar, and Yoshua Bengio. Learning to navigate the synthetically accessible chemical space using reinforcement learning. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3668–3679. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/gottipati20a.html.

Christoph Grebner, Erik Malmerberg, Andrew Shewmaker, Jose Batista, Anthony Nicholls, and Jens Sadowski. Virtual screening in the cloud: How big is big enough? *Journal of Chemical Information and Modeling*, 60(9):4274–4282, 2020. URL https://doi.org/10.1021/acs.jcim.9b00779.

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. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=kJqXEPXMsE0.

Jeff Guo and Philippe Schwaller. Directly optimizing for synthesizability in generative molecular design using retrosynthesis models. *Chemical Science*, 16:6943–6956, 2025. URL https://doi.org/10.1039/d5sc01476j.

A. Gupta, A. T. Müller, B. J. H. Huisman, J. A. Fuchs, P. Schneider, and G. Schneider. Generative recurrent networks for de novo drug design. *Molecular Informatics*, 37(1–2):1700111, 2018. URL https://doi.org/10.1002/minf.201700111.

Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamin Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018. URL https://doi.org/10.1021/acscentsci.7b00572.

Markus Hartenfeller, Heiko Zettl, Miriam Walter, Matthias Rupp, Felix Reisen, Ewgenij Proschak, Sascha Weggen, Holger Stark, and Gisbert Schneider. DOGS: Reaction-driven de novo design of bioactive compounds. *PLoS Computational Biology*, 8(2):e1002380, 2012. URL https://doi.org/10.1371/journal.pcbi.1002380.

Paul C. D. Hawkins, A. Geoffrey Skillman, and Anthony Nicholls. Comparison of shape-matching and docking as virtual screening tools. *Journal of Medicinal Chemistry*, 50(1):74–82, 2007. URL https://doi.org/10.1021/jm0603365.

Paul C. D. Hawkins, A. Geoffrey Skillman, Gregory L. Warren, Benjamin A. Ellingson, and Matthew T. Stahl. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of Chemical Information and Modeling*, 50(4):572–584, 2010. URL https://doi.org/10.1021/ci100031x.

Irina Higgins, Loïc Matthey, Arka Pal, Christopher P. Burgess, Xavier Glorot, Matthew M. Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=Sy2fzU9gl.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. URL https://openreview.net/forum?id=qw8AKxfYbI.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Julien Horwood and Emmanuel Noutahi. Molecular design in synthetically accessible chemical space via deep reinforcement learning. *ACS Omega*, 5(51):32984–32994, 2020. URL https://doi.org/10.1021/acsomega.0c04153.

Zehuan Huang, Yuan-Chen Guo, Xingqiao An, Yunhan Yang, Yangguang Li, Zi-Xin Zou, Ding Liang, Xihui Liu, Yan-Pei Cao, and Lu Sheng. MIDI: Multi-instance diffusion for single image to 3D scene generation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 23646–23657, June 2025. URL https://openaccess.thecvf.com/content/CVPR2025/papers/Huang_MIDI_Multi-Instance_Diffusion_for_Single_Image_to_3D_Scene_Generation_CVPR_2025_paper.pdf.

Fergus Imrie, Thomas E. Hadfield, Anthony R. Bradley, and Charlotte M. Deane. Deep generative design with 3D pharmacophoric constraints. *Chemical Science*, 12:14577–14589, 2021. URL http://dx.doi.org/10.1039/D1SC02436A.

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, 2022. URL https://doi.org/10.1088/2632-2153/ac3ffb.

Ross Irwin, Alessandro Tibo, Jon Paul Janet, and Simon Olsson. SemlaFlow – Efficient 3D molecular generation with latent attention and equivariant flow matching. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025. URL https://openreview.net/forum?id=bee2G6pEh0.

Jan H. Jensen. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. *Chemical Science*, 10:3567–3572, 2019. URL http://dx.doi.org/10.1039/C8SC05372C.

Wengong Jin, Connor Coley, Regina Barzilay, and Tommi Jaakkola. Predicting organic reaction outcomes with Weisfeiler-Lehman network. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/ced556cd9f9c0c8315cfbe0744a3baf0-Paper.pdf.

Chaitanya K. Joshi, Xiang Fu, Yi-Lun Liao, Vahe Gharakhanyan, Benjamin Kurt Miller, Anuroop Sriram, and Zachary W. Ulissi. All-atom diffusion transformers: Unified generative modelling of molecules and materials. In *International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=89QPmZjIhv.

Joonyoung F Joung, Mun Hong Fong, Nicholas Casetti, Jordan P Liles, Ne S Dassanayake, and Connor W Coley. Electron flow matching for generative reaction mechanism prediction. *Nature*, 645(8079):115–123, 2025. URL https://doi.org/10.1038/s41586-025-09426-9.

Matthew A Kayala, Chloé-Agathe Azencott, Jonathan H Chen, and Pierre Baldi. Learning to predict chemical reactions. *Journal of Chemical Information and Modeling*, 51(9):2209–2222, 2011. URL https://doi.org/10.1021/ci200207y.

Diederik P Kingma and Max Welling. Auto-encoding variational Bayes, 2013. URL https://arxiv.org/abs/1312.6114.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. ChemBO: Bayesian optimization of small organic molecules with synthesizable recommendations. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 3393–3403. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/korovina20a.html.

MichałKoziarski, Andrei Rekesh, Dmytro Shevchuk, Almer van der Sloot, Piotr Gaiński, Yoshua Bengio, Cheng-Hao Liu, Mike Tyers, and Robert A. Batey. RGFN: Synthesizable molecular generation using GFlowNets. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 46908–46955. Curran Associates, Inc., 2024. URL https://doi.org/10.52202/079017-1488.

Jacqueline Kuan, Mariia Radaeva, Adeline Avenido, Artem Cherkasov, and Francesco Gentile. Keeping pace with the explosive growth of chemical libraries with structure-based virtual screening. *WIREs Computational Molecular Science*, 13(6):e1678, 2023. URL https://doi.org/10.1002/wcms.1678.

Tuan Le, Julian Cremer, Frank Noe, Djork-Arné Clevert, and Kristof T Schütt. Navigating the design space of equivariant diffusion-based generative models for de novo 3D molecule generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=kzGuiRXZrQ.

Seul Lee, Karsten Kreis, Srimukh Prasad Veccham, Meng Liu, Danny Reidenbach, Saee Paliwal, Weili Nie, and Arash Vahdat. Rethinking molecule synthesizability with chain-of-reaction, 2025. URL https://arxiv.org/abs/2509.16084.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics. URL https://aclanthology.org/2020.acl-main.703/.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, and Vijay Pande. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Central Science*, 3(10):1103–1113, 2017. URL https://doi.org/10.1021/acscentsci.7b00303.

Cheng-Hao Liu, Maksym Korablyov, Stanisław Jastrzebski, Paweł Włodarczyk-Pruszyński, Yoshua Bengio, and Marwin Segler. RetroGNN: Fast estimation of synthesizability for virtual screening and de novo design by learning from slow retrosynthesis software. *Journal of Chemical Information and Modeling*, 62(10):2293–2300, 2022. URL https://doi.org/10.1021/acs.jcim.1c01476.

Alston Lo, Connor W Coley, and Wojciech Matusik. A genetic algorithm for navigating synthesizable molecular spaces, 2025. URL https://arxiv.org/abs/2509.20719.

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(1):20, 2024. URL https://doi.org/10.1186/s13321-024-00812-5.

Daniel Lowe. Chemical reactions from US patents (1976-Sep2016). Jun 2017. URL https://doi.org/10.6084/m9.figshare.5104873.v1.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. A 3D generative model for structure-based drug design. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 6229–6239. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/314450613369e0ee72d0da7f6fee773c-Paper.pdf.

Shitong Luo, Wenhao Gao, Zuofan Wu, Jian Peng, Connor W. Coley, and Jianzhu Ma. Projecting molecules into synthesizable chemical spaces. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 33289–33304. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/luo24a.html.

Omar Mahmood, Pedro O. Pinheiro, Richard Bonneau, Saeed Saremi, and Vishnu Sresht. Pharmacophore-based design by learning on voxel grids, 2025. URL https://arxiv.org/abs/2512.02031.

Krzysztof Maziarz, Guoqing Liu, Austin Tripp, Junren Li, Piotr Gaiński, and Marwin Segler. Chemist-aligned retrosynthesis by ensembling diverse inductive bias models. In *NeurIPS 2025 AI for Science Workshop*, 2025. URL https://openreview.net/forum?id=F05pziBqOg.

David Mendez, Anna Gaulton, A Patrícia Bento, Jon Chambers, Marleen De Veij, Eloy Félix, María Paula Magariños, Juan F Mosquera, Prudence Mutowo, Michał Nowotka, María Gordillo-Marañón, Fiona Hunter, Laura Junco, Grace Mugumbate, Milagros Rodriguez-Lopez, Francis

Atkinson, Nicolas Bosc, Chris J Radoux, Aldo Segura-Cabrera, Anne Hersey, and Andrew R Leach. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Research*, 47(D1): D930–D940, 2019. URL https://doi.org/10.1093/nar/gky1075.

Juno Nam and Jurae Kim. Linking the neural machine translation and the prediction of organic chemistry reactions, 2016. URL https://arxiv.org/abs/1612.09529.

NextMove Software. Pistachio: Reaction data, querying and analytics. https://www.nextmovesoftware.com/pistachio.html, 2025. Accessed 2025-09-11.

Pedro O O. Pinheiro, Joshua Rackers, Joseph Kleinhenz, Michael Maser, Omar Mahmood, Andrew Watkins, Stephen Ra, Vishnu Sresht, and Saeed Saremi. 3D molecule generation by denoising voxel grids. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 69077–69097. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/da1131a86ac3c70e0b7cae89c3d4df22-Paper-Conference.pdf.

Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1):48, 2017. URL https://doi.org/10.1186/s13321-017-0235-x.

Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013. URL https://doi.org/10.1016/j.commatsci.2012.10.028.

Kostas Papadopoulos, Kathryn A. Giblin, Jon Paul Janet, Atanas Patronov, and Ola Engkvist. De novo design with deep generative models based on 3D similarity scoring. *Bioorganic & Medicinal Chemistry*, 44:116308, 2021. URL https://doi.org/10.1016/j.bmc.2021.116308.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182, Los Alamitos, CA, USA, October 2023. IEEE Computer Society. URL https://doi.ieeecomputersociety.org/10.1109/ICCV51070.2023.00387.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. Pocket2Mol: Efficient molecular sampling based on 3D protein pockets. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 17644–17655. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/peng22b.html.

Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, Artur Kadurin, Simon Johansson, Hongming Chen, Sergey Nikolenko, Alán Aspuru-Guzik, and Alex Zhavoronkov. Molecular Sets (MOSES): A benchmarking platform for molecular generation models. *Frontiers in Pharmacology*, 11, 2020. URL https://doi.org/10.3389/fphar.2020.565644.

Kristina Preuer, Philipp Renz, Thomas Unterthiner, Sepp Hochreiter, and Günter Klambauer. Fréchet ChemNet Distance: A metric for generative models for molecules in drug discovery. *Journal of Chemical Information and Modeling*, 58(9):1736–1741, 2018. URL https://doi.org/10.1021/acs.jcim.8b00234.

Yiming QIN, Manuel Madeira, Dorina Thanou, and Pascal Frossard. DeFoG: Discrete flow matching for graph generation. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=KPRIwWhqAZ.

Andrei Rekesh, Miruna Cretu, Dmytro Shevchuk, Vignesh Ram Somnath, Pietro Liò, Robert A. Batey, Mike Tyers, Michał Koziarski, and Cheng-Hao Liu. SynCoGen: Synthesizable 3D molecule generation via joint reaction and coordinate modeling, 2025. URL https://arxiv.org/abs/2507.11818.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/rezende15.html.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, Los Alamitos, CA, USA, June 2022. IEEE Computer Society. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042.

Mikołaj Sacha, Mikołaj Błaz, Piotr Byrski, Paweł Dabrowski-Tumanski, Mikołaj Chrominski, Rafał Loska, Paweł Włodarczyk-Pruszynski, and Stanisław Jastrzebski. Molecule edit graph attention network: Modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling*, 61(7):3273–3284, 2021. URL https://doi.org/10.1021/acs.jcim.1c00537.

Tatsuya Sagawa and Ryosuke Kojima. ReactionT5: a pre-trained transformer model for accurate chemical reaction prediction with limited data. *Journal of Cheminformatics*, 17(1):126, 2025. URL https://doi.org/10.1186/s13321-025-01075-4.

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. URL https://doi.org/10.1038/s43588-024-00737-x.

Arne Schneuing, Ilia Igashov, Adrian W. Dobbelstein, Thomas Castiglione, Michael M. Bronstein, and Bruno Correia. Multi-domain distribution learning for de novo drug design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=g3VCIM94ke.

Philippe Schwaller, Theophile Gaudin, David Lanyi, Costas Bekas, and Teodoro Laino. "Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chemical Science*, 9(28):6091–6098, 2018. URL https://doi.org/10.1039/C8SC02339E.

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. URL https://doi.org/10.1021/acscentsci.9b00576.

Marwin HS Segler, Thierry Kogej, Christian Tyrchan, and Mark P Waller. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Central Science*, 4(1): 120–131, 2018. URL https://doi.org/10.1021/acscentsci.7b00512.

Philipp Seidl, Philipp Renz, Natalia Dyubankova, Paulo Neves, Jonas Verhoeven, Jorg K Wegner, Marwin Segler, Sepp Hochreiter, and Gunter Klambauer. Improving few-and zero-shot reaction template prediction using modern hopfield networks. *Journal of Chemical Information and Modeling*, 62(9):2111–2120, 2022. URL https://doi.org/10.1021/acs.jcim.1c01065.

Seonghwan Seo, Minsu Kim, Tony Shen, Martin Ester, Jinkyoo Park, Sungsoo Ahn, and Woo Youn Kim. Generative flows on synthetic pathway for drug design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=pB1XSj2y4X.

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. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=4aXfSLfM0Z.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/3001ef257407d5a371a96dcd947c7d93-Paper.pdf.

Megan Stanley and Marwin Segler. Fake it until you make it? Generative de novo design and virtual screening of synthesizable molecules. *Current Opinion in Structural Biology*, 82:102658, 2023. URL https://doi.org/10.1016/j.sbi.2023.102658.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger (eds.), *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5a18e133cbf9f257297f410bb7eca942-Paper.pdf.

Kyle Swanson, Gary Liu, Denise B. Catacutan, Autumn Arnold, James Zou, and Jonathan M. Stokes. Generative AI for designing and validating easily synthesizable and structurally novel antibiotics. *Nature Machine Intelligence*, 6(3):338–353, 2024. URL https://doi.org/10.1038/s42256-024-00809-7.

Kyle Swanson, Gary Liu, Denise B. Catacutan, Stewart McLellan, Autumn Arnold, Megan M. Tu, Eric D. Brown, James Zou, and Jonathan M. Stokes. SyntheMol-RL: a flexible reinforcement learning framework for designing novel and synthesizable antibiotics. *bioRxiv*, 2025. URL https://doi.org/10.1101/2025.05.17.654017.

Viet-Khoa Tran-Nguyen, Célien Jacquemard, and Didier Rognan. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *Journal of Chemical Information and Modeling*, 60 (9):4263–4273, 2020. URL https://doi.org/10.1021/acs.jcim.0c00155.

Austin Tripp, Erik Daxberger, and José Miguel Hernández-Lobato. Sample-efficient optimization in the latent space of deep generative models via weighted retraining. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 11259–11272. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/81e3225c6ad49623167a4309eb4b2e75-Paper.pdf.

Zhengkai Tu and Connor W Coley. Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *Journal of Chemical Information and Modeling*, 62 (15):3503–3513, 2022. URL https://doi.org/10.1021/acs.jcim.2c00321.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 11287–11302. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/5dca4c6b9e244d24a30b4c45601d9720-Paper.pdf.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https:

//proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

H. Maarten Vinkers, Marc R. de Jonge, Frederik F. D. Daeyaert, Jan Heeres, Lucien M. H. Koymans, Joop H. van Lenthe, Paul J. Lewi, Henk Timmerman, Koen Van Aken, and Paul A. J. Janssen. SYNOPSIS: SYNthesize and OPtimize system in silico. *Journal of Medicinal Chemistry*, 46(13): 2765–2773, 2003. URL https://doi.org/10.1021/jm030809x.

Carlos Vonessen, Charles Harris, Miruna Cretu, and Pietro Liò. Tabasco: A fast, simplified model for molecular generation with improved physical quality, 2025. URL https://arxiv.org/abs/2507.00899.

Jennifer N Wei, David Duvenaud, and Alán Aspuru-Guzik. Neural networks for the prediction of organic chemistry reactions. *ACS Central Science*, 2(10):725–732, 2016. URL https://doi.org/10.1021/acscentsci.6b00219.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In Qun Liu and David Schlangen (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Online, October 2020. Association for Computational Linguistics. URL https://aclanthology.org/2020.emnlp-demos.6/.

Weixin Xie, Jianhang Zhang, Qin Xie, Chaojun Gong, Yuhao Ren, Jin Xie, Qi Sun, Youjun Xu, Luhua Lai, and Jianfeng Pei. Accelerating discovery of bioactive ligands with pharmacophore-informed generative models. *Nature Communications*, 16(1):2391, 2025. URL https://doi.org/10.1038/s41467-025-56349-0.

Minkai Xu, Alexander Powers, Ron Dror, Stefano Ermon, and Jure Leskovec. Geometric latent diffusion models for 3D molecule generation, 2023. URL https://arxiv.org/abs/2305.01140.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023. URL https://dl.acm.org/doi/10.5555/3618408.3620080.

Kevin Yu, Jihye Roh, Ziang Li, Wenhao Gao, Runzhong Wang, and Connor W. Coley. Double-ended synthesis planning with goal-constrained bidirectional search. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 112919–112949. Curran Associates, Inc., 2024. URL https://doi.org/10.52202/079017-3588.

Huimin Zhu, Renyi Zhou, Dongsheng Cao, Jing Tang, and Min Li. A pharmacophore-guided deep learning approach for bioactive molecular generation. *Nature Communications*, 14(1):6234, 2023. URL https://doi.org/10.1038/s41467-023-41454-9.

Yael Ziv, Fergus Imrie, Brian Marsden, and Charlotte M. Deane. MolSnapper: Conditioning diffusion for structure-based drug design. *Journal of Chemical Information and Modeling*, 65(9): 4263–4273, 2025. URL https://doi.org/10.1021/acs.jcim.4c02008.

A Appendix

B Background and related work

Structure- and ligand-based drug design. Many generative models have been proposed to sample small molecules in sequence (Gupta et al., 2018; Blaschke et al., 2020; Segler et al., 2018), 2D (Jensen, 2019; Jin et al., 2017; QIN et al., 2025), 3D (Le et al., 2024; Dunn & Koes, 2024; Huang

et al., 2025; Irwin et al., 2025; Vonessen et al., 2025), and voxel spaces (O. Pinheiro et al., 2023). Conditional generation for protein binders can be broadly categorized into pocket-conditioned approaches, which condition on binding-site representations and implicitly capture protein-ligand interactions, and interaction-conditioned approaches, which condition directly on desired interaction cues such as 3D shape, pharmacophores, or electrostatic potential surfaces. The latter is most relevant to our work. Several methods generate SMILES strings or molecular graphs conditioned on 3D pharmacophore representations (Imrie et al., 2021; Zhu et al., 2023; Xie et al., 2025; Mahmood et al., 2025). MolSnapper (Ziv et al., 2025) and, more recently, ShEPhERD (Adams et al., 2025) instead condition on richer interaction profiles and generate molecules directly in 3D space. However, MolSnapper supports a limited set of pharmacophore types, and neither method explicitly constrains generation to synthesizable chemical space.

Synthesis-aware generation. Parallel work has focused on enhancing synthesizability by generating molecules via reaction pathways. Composing these reaction pathways into full synthesis plans is a challenging discrete search problem, and existing methods span diverse paradigms, including reinforcement learning (Gottipati et al., 2020; Horwood & Noutahi, 2020), search (Swanson et al., 2025), surrogate-guided optimization (Korovina et al., 2020), genetic algorithms (Lo et al., 2025), and GFlowNets (Cretu et al., 2025; Koziarski et al., 2024; Seo et al., 2025). One particularly effective strategy is to cast synthesis plans as sequences and learn *serialized* pathway representations autoregressively (Bradshaw et al., 2020; Gao et al., 2022; Lee et al., 2025; Luo et al., 2024; Gao et al., 2025), analogous to next-token prediction in language modeling (Vaswani et al., 2017; Bengio et al., 2003; Sutskever et al., 2014). By amortizing the cost of discrete search, such models learn mappings from continuous representations to synthesis plans and can be integrated into larger frameworks (e.g., encoder-decoder architectures) for tasks such as molecular optimization or retrosynthesis (i.e., predicting a route for a given target molecule). While these approaches enforce viable synthesis routes, they do not directly control final 3D geometry. Recently, CGFlow (Shen et al., 2025) and SynCoGen (Rekesh et al., 2025) jointly design synthesis pathways and 3D poses, but they rely on fixed reaction template libraries and work in ambient space.

Latent diffusion models. Latent diffusion models (Vahdat et al., 2021; Rombach et al., 2022) perform diffusion (Sohl-Dickstein et al., 2015; Song & Ermon, 2019) in the learned latent space of an autoencoder rather than directly in the high-dimensional input space, enabling more efficient training and sampling. This paradigm has been highly effective in image, audio, and video generation, especially when combined with Diffusion Transformers (DiTs) (Peebles & Xie, 2023), which show that standard transformer backbones scale effectively as denoisers. In molecular modeling, Xu et al. (2023) introduced latent diffusion in the space of an equivariant autoencoder, while Joshi et al. (2025) proposed a unified (non-equivariant) latent diffusion model for small molecules and periodic materials. However, these approaches do not demonstrate latent generation conditioned on rich, chemically grounded features. We address this gap by enabling pharmacophore-conditioned latent generation and augmenting the VAE with an auxiliary synthesis-decoder head that produces explicit reaction pathways, jointly targeting native-3D conditional generation and synthesizability.

C METHODS

C.1 DATASET

For our dataset, we make use of the provided synthesis DAGs from Bradshaw et al. (2020). Briefly, these DAGs were obtained from a cleaned version of the USPTO reaction dataset after stripping out reagents (Jin et al., 2017; Lowe, 2017); first, by building up a reaction network from the USPTO reactions and an initial set of "building block" nodes (picked by frequency), before extracting a possible synthesis plan for each non-building block node by tracing reactions backward to a loop-free subgraph that terminates in building blocks (selecting a single route when multiple alternatives exist). Using this data, we first filter the reactions such that the final products contain between 10 and 40 heavy atoms, and obtain $67\,512$ synthesis DAGs, 90% of which we keep as our training set and use the remaining data for our validation and test sets. Note that we randomly split by final molecule (i.e., the different train, validation, and test sets lead to different final molecules but may share reactions/sub-networks); analyzing SynLaD's ability to extrapolate to different data splits (for instance, a "Reachable" vs "Hard" split akin to Yu et al., 2024, p.7) is an interesting future direction

to explore. For each final product of the synthesis DAGs we compute a maximum of 10 low-energy conformers using Omega (Hawkins et al., 2010), which results in a total of $584\,896$ 3D molecules.

C.2 SynLaD: model design

We hypothesize that the two-stage learning process of SynLaD offers two advantages. First, the shared latent space aligns the two modalities—molecular conformations and synthesis pathways—by requiring both decoders to reconstruct the same molecule from a single latent representation. Second, as shown in Joshi et al. (2025), operating in a latent space reduces the complexity of generation and enables explicit pharmacophore conditioning by decoupling discrete and continuous spaces.

C.2.1 Multi-headed autoencoder for multi-modal reconstruction.

Our autoencoder consists of one encoder and two decoders trained using a combined reconstruction loss with a regularization term. We base our implementation of the 3D encoder and decoder on Joshi et al. (2025) and represent molecules using discrete atom types $\boldsymbol{A} = \{a_i\}_{i=1}^{N} \in \mathbb{Z}^{1\times N}$ and continuous coordinates $\boldsymbol{X} = \{x_i\}_{i=1}^{N} \in \mathbb{R}^{3\times N}$, where N is the number of atoms. Given a molecule, the encoder f projects $\boldsymbol{A}$ and $\boldsymbol{X}$ into a per-atom latent representation $\boldsymbol{Z} = f(\boldsymbol{A}, \boldsymbol{X})$, and the 3D decoder g_{3D} reconstructs the molecule from the latent, giving $\boldsymbol{A'}, \boldsymbol{X'} = g_{\text{3D}}(\boldsymbol{Z})$, where $\boldsymbol{Z} = \{z_i\}_{i=1}^{N} \in \mathbb{R}^{d\times N}$. We use a standard transformer (Vaswani et al., 2017) and learn molecular symmetries via random rotations.

Synthesis decoder We define an auxiliary decoder $g_{\text{syn}}(\boldsymbol{Z})$, which, given a latent $\boldsymbol{Z}$, generates a synthesis pathway for the molecule represented in 3D by $(\boldsymbol{A}, \boldsymbol{X}) = g_{\text{3D}}(\boldsymbol{Z})$. A synthesis pathway can be represented using a directed acyclic graph (DAG), where nodes represent molecules and edges represent reactions (see Figure 1). Such a DAG shows how parent building block molecules (source nodes) react to first form more complex intermediate products and eventually a final product (a single sink node). To model such a DAG we can *serialize* it, in a bottom-up manner similar to Bradshaw et al. (2020), into a linear sequence of tokens following the recipe in Appendix C.3. Once we have such a synthesis sequence $\mathbf{S} = (s_1, \ldots, s_L)$, we train a causal transformer model to generate it by autoregressively modeling the conditional probability $p(s_i|\boldsymbol{S}_{<\boldsymbol{i}}, \boldsymbol{Z})$. We embed each token in the sequence using a custom embedding scheme (see Appendix C.4) and train the transformer using teacher forcing conditioned with cross-attention on the latents.

Product identity In our serialized synthesis plan representation, the identity of each product node is determined by the identities of the corresponding reactant molecules. During training, we use a dataset containing full synthesis plans (see Section 3), allowing the correct reaction products to be inserted directly. At test time, however, product information is not available and must be inferred. We therefore employ a reaction-prediction oracle that can predict the most likely product given a set of reactants. The oracle is a transformer-based encoder-decoder model based on the BART architecture (Lewis et al., 2020), closely related to the Molecular Transformer of Schwaller et al. (2019). Details of its training are provided in Appendix C.9.

C.2.2 Conditional generative modeling of latent representations

We train a flow matching model (Lipman et al., 2023; Song & Ermon, 2019; Ho et al., 2020; Joshi et al., 2025) on the latent space learned by our autoencoder, by denoising latent samples from a base Gaussian distribution towards a target distribution represented by latents of ground-truth molecules.

Pharmacophore embeddings. A pharmacophore abstracts components of a molecule into predefined key interaction features, providing a compact representation of interaction-relevant moieties and their spatial coordinates. We represent pharmacophores using six discrete pharmacophore types $\boldsymbol{P} = \{p_i\}_{i=1}^{N_p} \in \mathbb{Z}^{1\times N_p}$ (hydrogen bond donor, hydrogen bond acceptor, cation, anion, aromatic ring and hydrophobe) and coordinates for each pharmacophore feature present in the molecule: $\boldsymbol{X}^{\text{ph}} = \{x_i^{\text{ph}}\}_{i=1}^{N_p} \in \mathbb{R}^{3\times N_p}$. These are extracted from the molecule conformation through a deterministic mapping of the pharmacophore's atom positions. The features are embedded into a unified learned representation using a standard transformer, and this embedding serves as conditioning for

the diffusion model via cross-attention. The parameters of the pharmacophore transformer encoder and those of the DiT are jointly optimized via the conditional flow matching loss:

$$\mathcal{L}_{\text{CFM}} = \mathbb{E}_{(\boldsymbol{P},\boldsymbol{X}^{\text{ph}},\boldsymbol{Z}_1)\sim p_{\text{data}},\, t\sim\mathcal{U}[0,1],\, \boldsymbol{Z}_0\sim\mathcal{N}(0,I)}\left[\left\|u_\theta(\boldsymbol{Z}_t, t, \tau_\theta(\boldsymbol{P},\boldsymbol{X}^{\text{ph}})) - v_t(\boldsymbol{Z}_t \mid \boldsymbol{Z}_0, \boldsymbol{Z}_1)\right\|_2^2\right] \quad (1)$$

where u_θ is the denoiser backbone, which predicts the vector field based on the current state z_t, the time step t, and the pharmacophore conditional embedding $\tau_\theta(\boldsymbol{P}, \boldsymbol{X}^{\text{ph}})$. v_t is the ground-truth vector field and the state $\boldsymbol{Z}_t = (1-t)\boldsymbol{Z}_0 + t\boldsymbol{Z}_1$ is a linear interpolation between a clean latent sample $\boldsymbol{Z}_1$ and a noise sample $\boldsymbol{Z}_0$ drawn from a standard normal distribution $\mathcal{N}(0, I)$. The time step t is sampled uniformly from $\mathcal{U}(0, 1)$. Leveraging the equivalence between the velocity and endpoint flow matching formulations (Lipman et al., 2023), the model receives $\boldsymbol{Z}_t$ as input and predicts the terminal points of the trajectory during training. Following Joshi et al. (2025), we use a Diffusion Transformer (DiT) (Peebles & Xie, 2023) as our denoiser architecture with self-conditioning (Yim et al., 2023) and adaptive layer norm for time-step t modulation.

C.3 Reaction network serialization

A synthesis pathway here is defined as a directed acyclic graph (DAG), where nodes represent molecules (each unique molecule maps to a unique node) and the edges represent reactions. In order to describe these pathways using an autoregressive language model we need to serialize them into a linear sequence of tokens. Our proposed approach follows (Bradshaw et al., 2020) (alternative serialization schemes have also been proposed, Gao et al., 2025; Lee et al., 2025; Luo et al., 2024) and is shown in Figure 4. The scheme relies on two kinds of tokens: *action tokens* and *molecular tokens*. Action tokens define a new operation to take with respect to the graph and are generally followed by a series of molecular nodes that specify the arguments required to define the action's effect. There are three types of action tokens:

- `'B'` adds a building block node to the network (these are easily purchasable compounds and the token is followed by a single molecule token defining the identity of the chosen building block).
- `'F'` adds a reaction edge (and a product node if the molecule does not yet exist in the network) to our network through a single forward reaction. This action is followed by selecting one or more molecule nodes (representing the existing molecules in the network) to use as reactants, a sub-action `'F_r'` (which indicates the reaction should be run) and a final product node (followed by `'P'`) that represents the product of the reaction.
- `'S'` is the stop action, which halts the prediction.

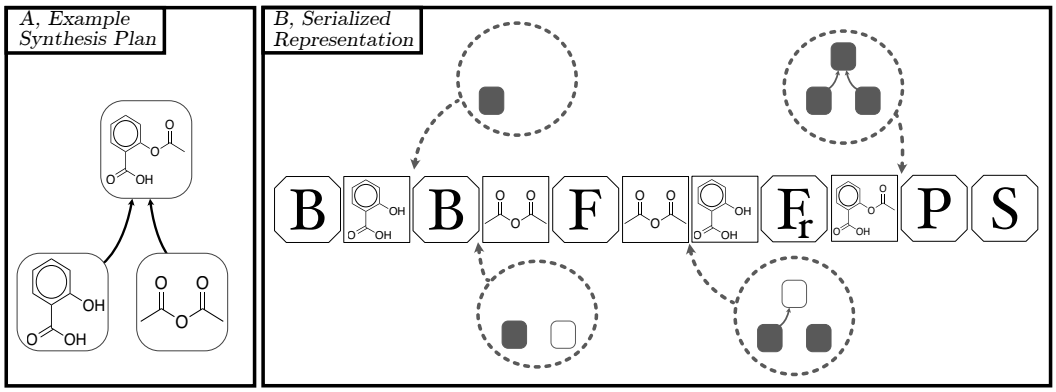

Figure 4: A tokenization scheme for describing synthesis plans in a bottom-up manner similar to Bradshaw et al. (2020). The synthesis plan, depicted as a DAG in **A** can be serialized as shown in **B** into a series of tokens. In **B**, the four dotted circles above the sequence indicate the state of the synthesis plan at that stage of decoding—empty nodes indicate the identity of the molecule corresponding to that node has yet to be defined. Note that this scheme can be used to describe complex, multi-step synthesis plans in a bottom-up manner.

Table 3: **Autoencoder loss weights ablation study.** We report 3D molecule match rate (computed with MoleculeMatcher from PyMatGen) and invariant building block match rate in the synthesis sequence.

λ	β	3D molecule match rate	Building block match rate
0.50	10^{-5}	82.1	83.3
0.70	10^{-5}	90.2	83.1
0.92	10^{-5}	98.5	82.5
0.98	10^{-5}	98.7	80.6
0.92	10^{-4}	80.6	67.4

Table 4: **Autoencoder reconstruction accuracies.** The synthesis match rate and Morgan fingerprint Tanimoto similarity are calculated between the final product of the predicted and the ground truth synthesis sequences. We report performance with two different inference strategies for the synthesis decoder.

Autoencoder hyperparameters		**3D metrics**		**Synthesis metrics**	
Inference method	Latent dim.	Match Rate (%)↑	RMSD (Å) ↓	Match Rate (%)↑	Tanimoto similarity↑
Beam search	8	98.9	0.05	58.4	0.82
Beam search	16	98.5	0.05	63.4	0.84
Beam search	32	98.9	0.04	62.7	0.83
Sampling	8	98.9	0.05	42.8	0.73
Sampling	16	98.5	0.05	49.0	0.76
Sampling	32	98.5	0.04	51.4	0.77

C.4 Synthesis action embeddings

The input tokens to the transformer are obtained via summing a series of embeddings as shown in Figure 1C: a learned embedding which specifies the token type, a molecule embedding (if the action involves a molecule) which is a projection of the molecule's 2048-bit Morgan fingerprint, and a learned positional embedding.

C.5 VAE training

Autoencoder ablations We report ablations for autoencoder hyperparameters in Tables 3 and 4. Increasing the weight on the synthesis decoder loss (see Eq. equation 2) harms 3D coordinate reconstruction and adds little benefit for the synthesis sequence reconstruction. Note that we report the accuracy of builiding block prediction in its order-invariant form (i.e., building blocks might be added earlier or later on in the sequence, with the end product being the same); this differs during training in which we randomly pick an order (for building blocks at the same level) and train on this fixed sequence using teacher forcing. We use $\lambda = 0.92$ in our experiments and $\beta = 10^{-5}$, as larger KL weight harms reconstruction accuracy for both decoders. We use a latent dimension $d = 16$ which shows the best molecule and synthesis reconstruction.

C.5.1 Synthesis decoder training

The synthesis decoder is implemented as a decoder-only transformer. Each input token is embedded by summing (i) a learned embedding for special tokens or retrieved molecular embeddings, (ii) a learned token-type embedding, and (iii) a learned positional embedding. Graph-level information is provided via a fingerprint encoder. The embedded sequence is processed by 12 transformer blocks with RMSNorm pre-normalization, multi-head self-attention, and a SwiGLU feed-forward network (further details can be found in our attached code). When conditioning is enabled, each block additionally applies a cross-attention sublayer over an external conditioning vector. During training,

we use teacher forcing (Goodfellow et al., 2016, §10.2.1), i.e., for each prediction step the decoder is conditioned on the ground-truth previous tokens. We compute the cross-entropy loss only on positions where the model is expected to generate tokens, and mask out product tokens (and any padded positions) so they do not contribute to the loss.

C.6 AUTOENCODER TRAINING OBJECTIVE

The two decoders are jointly trained with a weighted loss: the 3D decoder uses cross-entropy and MSE losses to reconstruct atom types and coordinates, while the synthesis decoder is optimized with a cross-entropy loss between its output logits and the ground-truth token sequence. We also introduce a per-channel KL-penalty:

$$\mathcal{L} = \lambda \mathcal{L}_{\text{3D}} + (1-\lambda)\mathcal{L}_{\text{synthesis}} + \beta\, D_{\text{KL}}(\mathcal{N}(\mu_{\mathbf{Z}}, \sigma_{\mathbf{Z}}) \parallel \mathcal{N}(0, I_d)) \tag{2}$$

where:

$$\mathcal{L}_{\text{3D}} = \frac{1}{N}\sum_{i=1}^{N} H(a_i, \hat{a}_i) + \frac{1}{3N}\sum_{i=1}^{N} \frac{\|\tilde{x}_i - \hat{\tilde{x}}_i\|^2}{\sigma^2}, \tag{3}$$

$$\mathcal{L}_{\text{synthesis}} = \frac{1}{L}\sum_{i=1}^{L} H(s_i, \hat{s}_i). \tag{4}$$

Here, $\tilde{x}$ represents zero-centered coordinates, $0 \leq \lambda \leq 1$ is the linear interpolation parameter between $\mathcal{L}_{\text{synthesis}}$ and $\mathcal{L}_{\text{3D}}$, L is the total sequence length, and s_i is the ground-truth synthesis token at step i. Following Higgins et al. (2017), Rombach et al. (2022), and Joshi et al. (2025) we introduce a per-channel KL-penalty weighted by β that regularizes the learned latent toward a standard normal distribution.

C.7 DIT TRAINING

We additionally introduce a cross-attention conditioning mechanism (trained with a conditioning dropout probability of 20%) to sample using classifier-free guidance during inference (Ho & Salimans, 2021):

$$\hat{u}_\theta(\boldsymbol{Z}_t, t, \tau_\theta(\boldsymbol{P}, \boldsymbol{X}^{\text{ph}})) = (1+w)u_\theta(\boldsymbol{Z}_t, t, \tau_\theta(\boldsymbol{P}, \boldsymbol{X}^{\text{ph}})) - w u_\theta(\boldsymbol{Z}_t, t, \emptyset) \tag{5}$$

where $\hat{u}_\theta$ is the final vector field prediction. This is a linear combination of the conditional and unconditional predictions, where $\emptyset$ represents the null token used for the dropped-out condition. The guidance scale w controls the strength of the conditioning signal and we experiment with both $w = 0$ and $w > 0$.

Sampling ablations. We investigate the effect of the number of sampling steps used during inference on molecule quality and synthesis/3D decoder output match rate in Table 5. Unless otherwise stated, we use 100 sampling steps in all experiments. We additionally ablate the classifier-free guidance weight w in Eq. equation 5 (Table 6). Increasing w improves control over latent quality and yields more valid decoded molecules, but reduces scaffold diversity. Since our primary objective is to maximize the number of generated hits, we therefore use $w = 0$ (i.e., no classifier-free guidance) throughout.

C.8 TRAINING AND HYPERPARAMETERS

We jointly train the VAE and synthesis decoder, then train the DiT in a second stage, using AdamW with a constant learning rate of 10^{-4}, zero weight decay, and a batch size of 256. We maintain an exponential moving average of the DiT parameters with decay 0.9999. Each stage is run to convergence, for up to 3000 epochs or 3 days on a single B200 GPU.

For the VAE, we use standard transformers for both the encoder and 3D decoder with hidden dimension $d_{\text{model}} = 256$, 4 attention heads and 6 layers (9.5M parameters total). For the synthesis decoder

Table 5: **Number of sampling steps ablation.** We carried out the ablation in the unconditional setting.

#Steps	Validity↑	PoseBusters↑	Synthesis/3D exact match rate↑
50	0.85	0.85	0.28
100	0.89	0.89	0.33
200	0.89	0.89	0.32
300	0.90	0.89	0.33
500	0.91	0.89	0.33

Table 6: **Classifier free guidance ablation.** We carried out the ablation for the Lit-PCBA ligands conditioning experiment and generated 100 samples per conditioning input.

cfg. setting	Validity↑	Diversity↑		Hits↑		Uniq. scaff. hits↑	
		(syn out)	(3D out)	(syn out)	(3D out)	(syn out)	(3D out)
w/o cfg.	0.49	0.86	0.86	17.9	38.1	6.9	18.8
$w = 0.5$	0.56	0.85	0.85	14.2	34.7	3.8	12.1
$w = 1.0$	0.46	0.85	0.84	14.3	38.0	4.8	15.2
$w = 2.0$	0.45	0.85	0.84	13.1	34.7	5.0	15.7
$w = 4.0$	0.36	0.85	0.84	11.7	23.4	5.2	12.6

we used $d_{\text{model}} = 512$, 16 attention heads, and 12 layers with causal self-attention, cross-attention conditioning to the latents, and SwigLU FFNs (65M parameters total).

For the second stage training, the pharmacophore encoder is a standard transformer with $d_{\text{model}} = 256$, 4 heads, and 4 layers (40M parameters total), and the DiT denoiser has $d_{\text{model}} = 768$, 12 attention heads, and 12 layers (130M parameters total).

C.9 Reaction predictor oracle training

As explained in the main text, when running our synthesis decoder at inference time we require access to a reaction predictor oracle to predict the molecular identity of the product molecule given the set of reactants (we ignore conditions and reagents in this current work). For this, we use a model based around the BART architecture (Lewis et al., 2020), using an implementation adapted from the `transformers` library (Wolf et al., 2020) for use in reaction prediction (Bradshaw et al., 2025). This is similar in spirit to other encoder-decoder models used for reaction prediction/one-step retrosynthesis (Nam & Kim, 2016; Schwaller et al., 2019; 2018; Liu et al., 2017). We train our reaction prediction oracle on reactant-product pairs from the same USPTO dataset we extract the synthesis plans from for training SynLaD; teacher forcing is used at training time and beam search is used during inference for making predictions. We leave to future work the exploration of using alternative reaction predictors (Tu & Coley, 2022; Jin et al., 2017; Bradshaw et al., 2019a; Seidl et al., 2022; Joung et al., 2025; Fooshee et al., 2018; Kayala et al., 2011; Sacha et al., 2021; Wei et al., 2016; Do et al., 2019; Bi et al., 2021; Irwin et al., 2022; Sagawa & Kojima, 2025), or even multiple at a time (Maziarz et al., 2025) for more robust predictions. We report accuracy results from training on a train/validation/test split of approximately 409k/ 30k/ 40k data points extracted from the USPTO dataset in Table 7.

C.10 Evaluation metrics

Metrics. We evaluate SynLaD on its ability to generate valid, novel, and realistic molecules. Sampled molecules generated by SynLaD are assessed using validity, internal diversity (Polykovskiy et al., 2020), and Fréchet ChemNet Distance (FCD) (Preuer et al., 2018). Diversity is further measured by the number of unique Bemis-Murcko scaffolds (Bemis & Murcko, 1996), defined as the molecule's ring systems and connecting linkers with all side chains removed. Synthesizability is additionally evaluated using AiZynthFinder (Genheden et al., 2020), and pose quality is assessed with

Table 7: **Reaction predictor test accuracy.** Top-k accuracy (%) on the held-out test set for product outcome prediction.

Checkpoint	Top-1	Top-2	Top-3	Top-5
Final (1M iters)	84.6	91.4	93.3	94.8

the PoseBusters benchmark (Buttenschoen et al., 2024) for molecules generated by our 3D decoder. For pharmacophore-conditioned experiments, we use OpenEye's ROCS scoring functions (Hawkins et al., 2007) to evaluate shape and pharmacophore overlap via Tanimoto Shape and Tanimoto Color scores, each ranging from 0 to 1 (with 0 representing no overlap and 1 representing perfect overlap), whose sum defines the Tanimoto Combo score. Following Grebner et al. (2020), we define a *hit* as a molecule with a Tanimoto Combo score of at least 1.2 relative to a query.

We further define the metrics used thoroughout the paper as follows:

- **Validity (Valid.)**: % of molecules that can be processed by RDKit.
- **Internal diversity (IntDiv.)**: One minus the value of the averaged pairwise Tanimoto similarities of the Morgan fingerprints (size 2048, radius 2) of the molecules in the generated set.
- **Murcko Scaffold count**: number of unique Murcko scaffolds in a generated set.
- **Novelty (Nov.)**: % of valid molecules not present in the training set.
- **PoseBusters (PB)**: % of molecules passing all PoseBusters filters (Buttenschoen et al., 2024). This was only evaluated for models that generated 3D poses.
- **FCD** (Fréchet ChemNet Distance): measures distributional similarity to a reference dataset using ChemNet embeddings (lower is better) (Preuer et al., 2018).
- **SA**: Averaged synthetic accessibility score from Ertl & Schuffenhauer (2009).
- **AiZynthFinder (AiZynth.)**: % of molecules (out of 100) for which AiZynthFinder finds valid synthetic pathways given its default set of reactions and ZINC building blocks.
- **RMSD and match rate**: For 3D molecule reconstruction we report RMSD and match rate computed using MoleculeMatcher from PyMatGen (Ong et al., 2013).
- **Hit rate**: We compute hit rate by counting the number of unique molecules with ROCS Tanimoto Combo score ≥ 1.2.

C.11 ROCS EVALUATION PIPELINE

We scored 3D similarity with OpenEye ROCS (Hawkins et al., 2007) using a standard OMEGA→ROCS workflow. We kept the query conformation intact (from our test set or Lit-PCBA benchmark pose) and for each candidate molecule, we first generated an ensemble of 100 low-energy 3D conformations with OMEGA (Hawkins et al., 2010), which constructs initial geometries from fragment conformers and then samples torsions using a knowledge-based torsion library, retaining a pruned set of energetically and geometrically reasonable conformers. We then used ROCS to perform fast rigid overlays between all query-candidate conformer pairs, representing molecular shape as a sum of atom-centered Gaussian functions and maximizing shared volume and pharmacophoric ("color") feature overlap. For each candidate, we reported the best overlay across conformers, and computed ShapeTanimoto, ColorTanimoto, and their sum (TanimotoCombo) as the final shape, color, and combined similarity scores. The same pipeline was applied for SynLaD and all baselines, regardless of the methods' output types.

D EXPERIMENTS

D.1 CONSISTENCY ANALYSIS: CROSS-DECODER AGREEMENT

To enable pharmacophore-conditioned sampling of synthesizable molecules that also match a desired 3D profile, we evaluate how well the two decoders remain aligned on *generated* samples.

Starting from a latent $\boldsymbol{Z}$ sampled from the trained diffusion model, we decode both a 3D molecule and a synthesis pathway and quantify how often these outputs correspond to the same, or closely matching, chemical features. We report cross-decoder agreement via exact identity and chemical/shape similarity in Table 5. The synthesis-decoded products preserve the 3D shape and pharmacophore characteristics implied by the 3D-decoded structures. We show examples of output pairs in Appendix 6.

Metric	Value
Exact match rate (%)	33.1
Scaffold match rate (%)	54.9

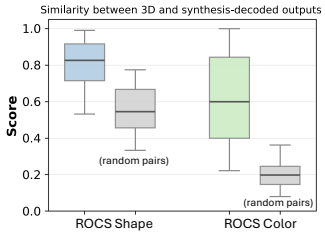

Figure 5: **Cross-decoder agreement on generated samples.** We evaluate similarities between synthesis- and 3D-decoded molecules obtained from the same latent. As a control, we also report ROCS Tanimoto Shape/Color scores for randomly paired synthesis- and 3D-decoded molecules, illustrating the similarity expected in the absence of cross-decoder coupling.

D.2 Pharmacophore conditioned generation

During training, we load the pharmacophore features for a conformer and randomly drop k features, where k is sampled uniformly from $[0, \max(0, N_{\text{ph}} - 3)]$, and N_{ph} denotes the number of pharmacophore features in the sample. During sampling, we set the latent to contain M atoms, with M sampled uniformly from $[\max(1, N - 3), N + 3]$, where N is the number of heavy atoms in the query (i.e., excluding hydrogens).

Table 8: **Molecule quality metrics for in-distribution pharmacophore conditioning.** Higher is better for all.

Method	Validity	Uniqueness	Diversity
SynLaD (3D out)	0.78	0.60	0.72
SynLaD (syn out)	1.0	0.63	0.75

D.3 Dataset screen

We report results for an experiment where we compare SynLaD to a dataset screen (see Section 3.3) in Figures 9 and 7.

Table 9: **Comparison to library screen.** Number of hits averaged over 10 queries from the test set. Here, we compare against the synthesis decoder of SynLaD as hits are more likely to be synthesizable and therefore able to be experimentally tested.

Method	Hits (avg.)
Dataset screen	49.7
SynLaD	59.4

D.4 Bioactive hit diversification experiment

We used ten PDB targets from Tran-Nguyen et al. (2020), with PDB IDs 4LDO, 5L2M, 2P15, 2IOK, 5FV7, 2V3D, 4ZZN, 6B73, 3ZME, 3A2I. We extracted the corresponding ligand for each of these as the 'query'. We show examples of decoded outputs, together with the conditioning query in Figure 9. We show additional results (including the per-query number of hits for the 3D decoder) in Figure 11.

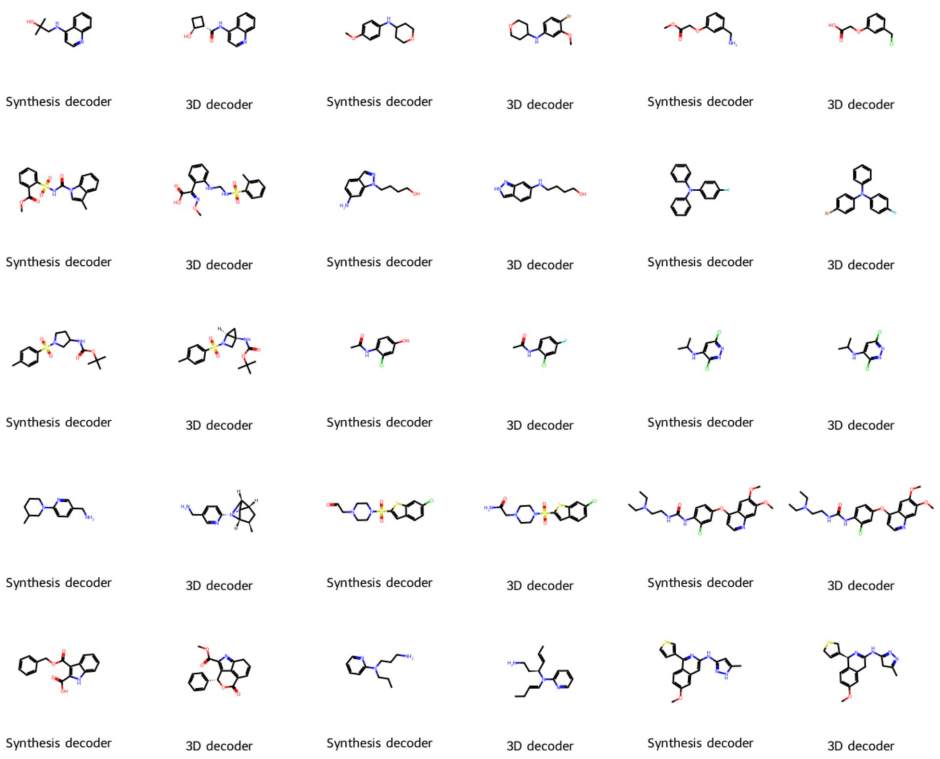

Figure 6: **Unconditional generation: synthesis- vs. 3D-decoded outputs.** Examples of decoded outputs from the synthesis decoder, followed by outputs of the 3D decoder, from the same latent.

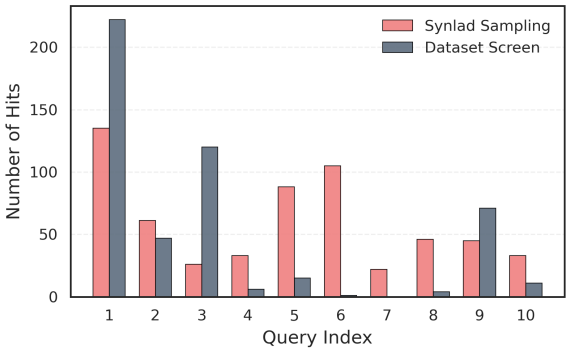

Figure 7: **Dataset screen vs. SynLaD sampling**. We report the number of unique hits for randomly selected queries from the test set. See the "Screening case study" section in the main text for further details.

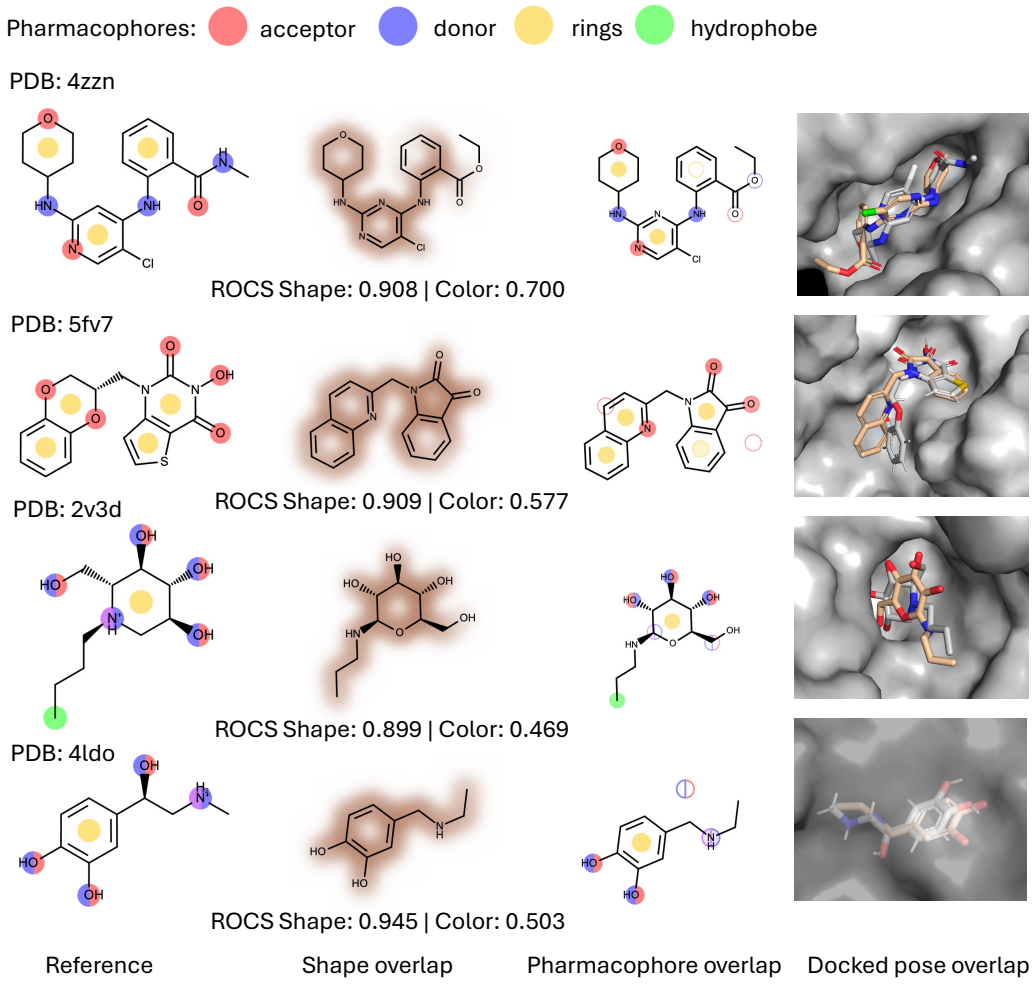

Figure 8: **Bioactive hit diversification experiment**. Examples of molecules generated by our model with annotated shape and pharmacophore overlap. The right-hand side shows the native ligand (gray) overlaid with the docked generated molecule (orange).

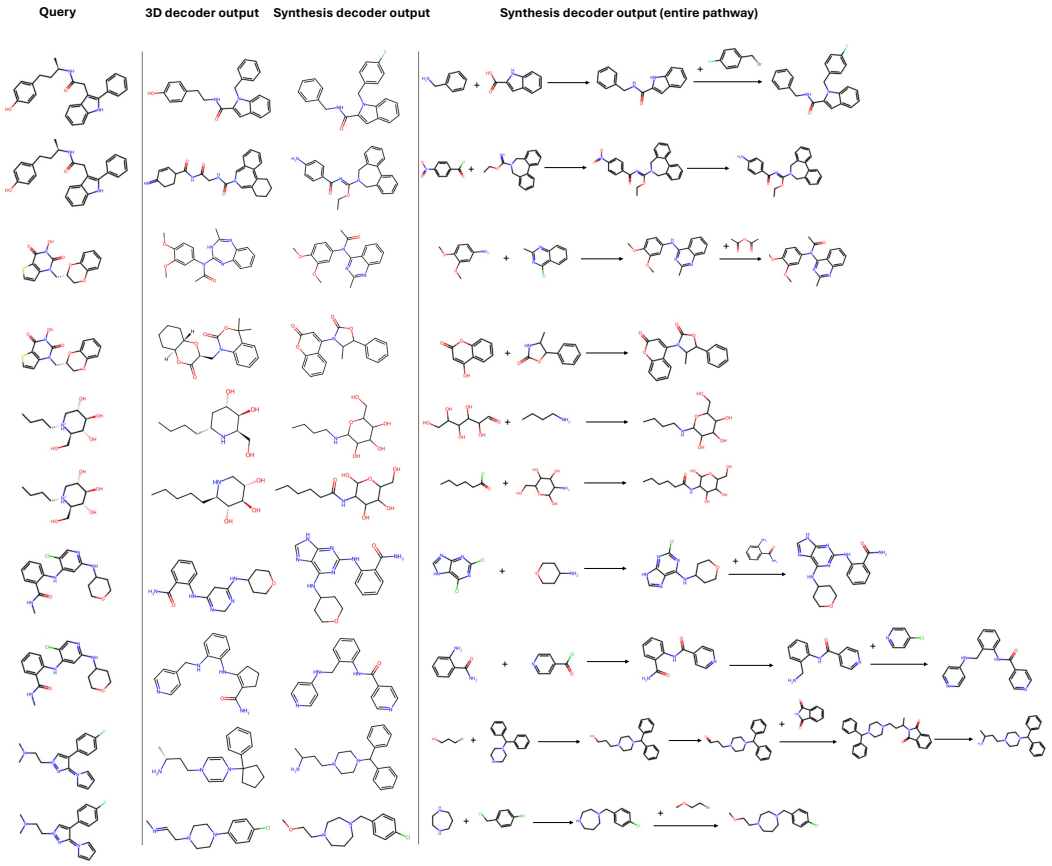

Figure 9: **SynLaD outputs for the Lit-PCBA conditioning task.** We visualize the Lit-PCBA ligand used as the conditioning **Query** (via its pharmacophore), together with the corresponding **3D decoder output** and the **synthesis decoder output** (final product and **entire pathway**). For readability, the query and 3D-decoded molecules are shown in 2D, although both are modeled in 3D coordinate space. We intentionally select examples where the synthesis- and 3D-decoder outputs do not coincide.

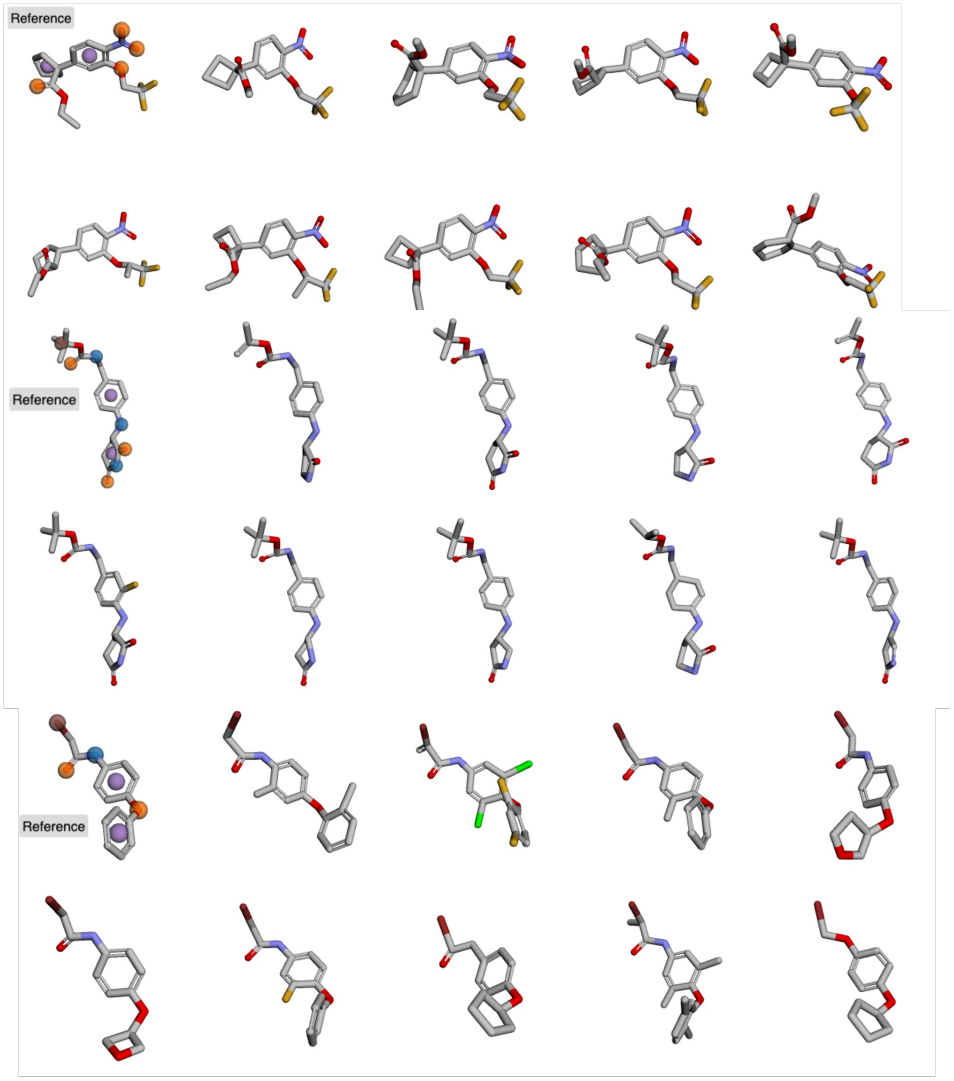

Figure 10: **Examples of query (reference) and sampled molecules conditioned on pharmacophores.** Samples were decoded using the 3D decoder, and we show the molecules in their predicted conformation.

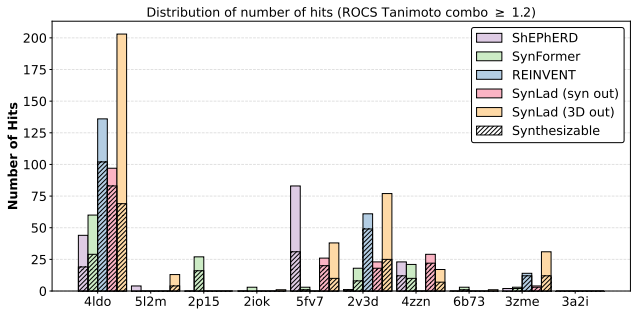

Figure 11: **Bioactive hit difersification experiment.** We add to the results in Figure 3 the performance of the 3D decoder.

D.5 Ablation: Removing the 3D branch

In order to see whether there is added benefit in modeling the conditioning input to the synthesis decoder using a latent diffusion framework, we perform an ablation where we remove the 3D latent component from SynLaD, and condition the synthesis decoder directly on the pharmacophore. We preserve the pharmacophore embedder described in Section C.2.2 and use the resulting embedding to condition on with an otherwise identical synthesis decoder via cross-attention. This yields a simplified pipeline, pharmacophore $\rightarrow$ pharmacophore embedding $\rightarrow$ synthesis pathway, which isolates the contribution of the 3D branch while keeping the synthesis decoder architecture and training procedure unchanged. We run the ablation on the in-distribution conditioning task reported in Section 3 and show in Table 10 that although the Tanimoto Combo scores do not suffer a drastic drop, the ablated version of the model appears to have collapsed to generating few samples per input (see Figure 12), showing that the entire framework is crucial to generating diverse hits.

Table 10: **Ablation: Removing the 3D branch.** We report metrics for outputs from SynLaD's synthesis decoder and an identical synthesis decoder conditioned directly on pharmacophores, without a latent space learned from 3D structure. The ablation was performed for the in-distribution experiment where we conditioned on 50 pharmacophore queries and generated 100 samples for each. Metrics are **averages** over all 50 conditioning inputs.

Method	**Div.** ↑	**Uniq.** ↑	**Tanimoto Combo** ↑	**#Hits** ↑	**#Scaff. hits** ↑
SynLaD (syn out)	0.75	0.63	1.18	27.5	11.6
w/o 3D latent	0.49	0.17	1.12	5.9	3.5

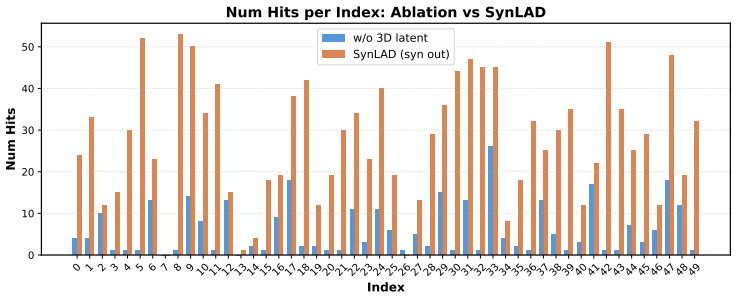

Figure 12: **Effect of removing the 3D branch on the number of hits.** We generate molecules conditioned on a pharmacophore with 1) SynLaD and a 2) synthesis decoder directly conditioned on the pharmacophore embedding.

D.6 COMPARISON TO BASELINES

ShEPhERD is an SE(3)-equivariant diffusion model that jointly denoises 3D molecular graphs and representations of their shapes and interaction profiles; we compare against their inpainting setting which conditions on electrostatic potential surfaces and pharmacophores. SynFormer is a transformer-based encoder-decoder framework that generates synthesizable 2D molecules in reaction-template and building-block space; we specifically compare against their SynFormerED variants for amortized analogue generation. As an orthogonal baseline, following Papadopoulos et al. (2021), we fine-tune REINVENT (Blaschke et al., 2020) using a ROCS Tanimoto color score as reward function. Unlike SynLaD and the other baselines, REINVENT does not perform amortized sampling and instead requires reinforcement learning–based optimization per query, making the comparison methodologically orthogonal (similar techniques can also be applied to autoencoders (Tripp et al., 2020)). REINVENT is also substantially more computationally expensive, requiring approximately 6 hours per query to generate the optimized samples here, compared to $\sim$1 min. for SynLaD, $\leq$1 min for SynFormer, and $\sim$28 min for ShEPhERD (Appendix D.6) per 100 samples, but we include it as it is a strong baseline in molecular generation.

ShEPhERD We used the official implementation at `https://github.com/coleygroup/shepherd` (Adams et al., 2025) and use the $p(x_1|x_3, x_4)$ conditional setting provided in the repository where x_1 denotes molecular structure, x_3 is the query electrostatic potential surface, and x_4 is the query's pharmacophore. We generate 500 analogues with number of atoms uniformly sampled from the interval $[\max(3, N-6), N+6]$, where N is the number of atoms in the query molecule (including hydrogens), to best match SynLaD's settings.

SynFormer We used the official implementation at `https://github.com/wenhao-gao/synformer` and changed the following inference settings to allow for higher quality designs compared to the default: `search_width=32, exhaustiveness=128, time_limit=300`. We generate and evaluate 500 molecules per query.

REINVENT We used the official implementation of REINVENT 4 available at `https://github.com/MolecularAI/REINVENT4` (Loeffler et al., 2024; Blaschke et al., 2020; Olivecrona et al., 2017). We start from the default REINVENT prior (from `https://zenodo.org/records/15641297`), which was trained on ChEMBL (Mendez et al., 2019) to cover a broad region of chemical space. We followed the repository's example reinforcement learning configuration (including hyperparameters), using a Tanimoto color scorer (with up to 5 conformers for computational speed) and a maximum of 300 steps (we do not include any other scoring components). We use the final 500 sampled molecules for evaluation. Although additional steps and hyperparameter tuning may further improve results, each target run already takes approximately 6 hours and can involve scoring up to 30 000 proposed molecules; therefore, we defer a comprehensive hyperparameter study (for both REINVENT and SynLaD) to future work.