Synthesizability via Reward Engineering: Expanding Generative Molecular AI into Synthetic Space

Dominik Dekleva, Jure Borišek *National Institute of Chemistry, Slovenia*
Martina H. Rambaher *Faculty of Pharmacy, University of Ljubljana, Slovenia*
Alexey Voronov, Hannes H. Loeffler, Jon Paul Janet *Molecular AI, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*
Albin Ekborg *Department of Physics, Chalmers University of Technology, Gothenburg, Sweden*

Abstract

Generating novel, drug-like molecules with realistic synthetic pathways is an essential goal in computer-aided drug discovery, yet generative models often lack synthesis awareness, resulting in compounds that are difficult or impossible to produce. To overcome this limitation, models must optimize not only molecular properties but also synthetic feasibility, which is not fully meaningful unless it incorporates user-defined factors like preferred reactions and available starting materials. Moreover, generating singleton compounds without respecting possibilities for parallel synthesis greatly increases the cost and complexity of synthesizing multiple proposed molecules. In practice, medicinal chemistry workflows group targets into families sharing coherent synthetic strategies and common intermediates, enabling efficient parallel and automated synthesis. Here we introduce SynthSense, a reinforcement learning framework that guides molecular design using retrosynthetic feedback. SynthSense offers extrinsic reward functions that assess molecule-level feasibility, such as adherence to available building blocks and preferred reactions, or synthesizability via predefined synthetic routes. It also implements intrinsic, batch-level functions that enforce route coherence across generated compounds. *In silico* multi-parameter validation demonstrated clear advantages over naïve approaches: SynthSense generated 6.2-fold more synthetically feasible hits than the control trained without SynthSense, achieved a 727-fold enrichment in hits synthesizable with a predefined synthetic route, and populated 2.3-fold more virtual parallel synthesis plates. These results demonstrate that by reframing synthesizability from a mere constraint into an active design objective, generative AI can better support the realities of modern medicinal chemistry by enabling personalized synthetic design, accelerating SAR exploration and aligning more naturally with automated parallel synthesis workflows.

Keywords: Generative AI, Drug Discovery, Synthetic Accessibility, Reinforcement Learning, Automated Synthesis

1 Introduction

Discovering novel, drug-like molecules is a cornerstone of pharmaceutical innovation, yet the scale of chemical space, estimated to contain over 10^{60} drug-like compounds [Polishchuk et al., 2013], makes exhaustive experimental exploration impossible. Deep generative models, including variational autoencoders [Oestreich et al., 2024], diffusion models [Oestreich et al., 2025], and reinforcement learning (RL)-guided frameworks [Olivecrona et al., 2017, Loeffler et al., 2024], have shown strong capabilities in producing chemically valid and property-optimized molecules. Yet many *in silico* designs are impossible to synthesize, limiting real-world impact.

Early approaches estimated synthetic accessibility via heuristics (e.g., SAscore, [Ertl and Schuffenhauer, 2009]) or machine learning predictors (e.g., SCScore, [Coley et al., 2018]). More recently, synthesis-conditioned generative models such as SynFlowNet [Cretu et al., 2024] have emerged, embedding synthesis biases during pretraining. While effective, these models hardwire synthetic constraints, making it difficult to adapt when project conditions shift, such as changes in reagent availability, synthetic preferences, or evolving goals. Retrosynthesis tools like AiZynthFinder [Loeffler et al., 2024] or ASKCOS [Coley et al., 2017], which predict plausible synthetic routes by recursively breaking down a molecule into purchasable or known precursors, are often applied *post hoc*, evaluating molecules after generation rather than guiding design. Current strategies therefore either constrain too early or intervene too late, failing to capture the dynamic reality of medicinal chemistry.

We propose **SynthSense**, a framework for synthesizability through reinforcement learning (RL) reward engineering. By shifting synthesis awareness to the post-training stage, SynthSense avoids the rigidity of synthesis-conditioned pretrained models and the passivity of *post hoc* retrosynthesis evaluation. This model-agnostic approach enables any pre-trained generator to adapt dynamically to reagent availability, preferred reactions, and evolving project goals without architectural modification or retraining. Critically, we argue that generative models should not design singleton molecules but rather molecular families with coherent synthetic strategies. In laboratory and robotic synthesis, grouping targets that share route archetypes enables parallel synthesis, reuse of intermediates, and higher throughput. SynthSense combines extrinsic (per-molecule) rewards with intrinsic (batch-level) rewards to align generation with practical synthetic considerations, expanding AI-driven molecular design from chemical space into *synthetic space*.

2 Methods

SynthSense leverages retrosynthesis information to guide molecular generation via RL rewards. In RL, molecular generation is treated as a sequential decision-making process: at each step, the generator constructs molecules token by token in SMILES format, producing a batch of compounds in each epoch. In each

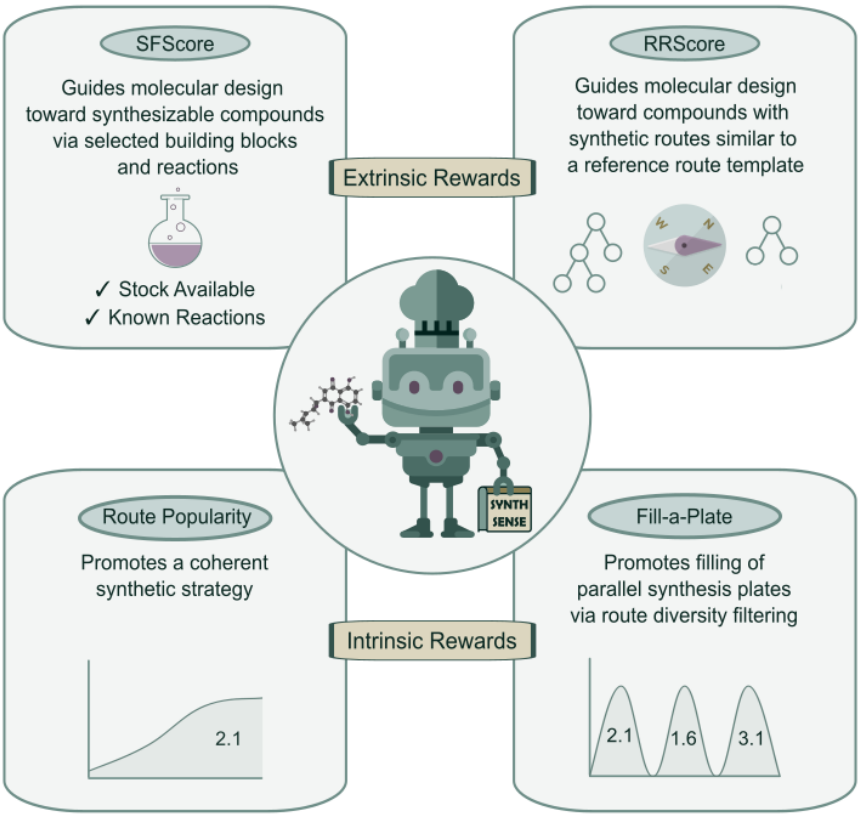

Figure 1: Overview of SynthSense reward functions.

batch, SynthSense evaluates individual molecules with extrinsic rewards and the batch as a whole with intrinsic rewards. These rewards are then used to update the generator's policy via policy gradient reinforcement learning (REINFORCE) algorithm [Williams, 1992], biasing future generations toward high-reward synthetically desirable molecules.

For each molecule in a batch, AiZynthFinder enumerates possible synthetic routes from available starting materials, and each reaction in a route is classified using NameRxn [NextMove, 2022]. Reward functions operate directly on these retrosynthetic tree representations rather than heuristic proxies.

Extrinsic rewards evaluate individual molecules:

- *Synthetic Feasibility Score (SFScore)* quantifies synthetic feasibility based on user-specified preferred reactions, starting materials, and route length. The highest-scoring molecule's tree defines the molecule's SFScore.
- *Reference Route Score (RRScore)* uses Tree Edit Distance (TED) [Pawlik

and Augsten, 2015, 2016] to bias molecules toward a specified synthetic pathway.

Intrinsic rewards operate at the batch level:

- *Route Popularity* promotes molecules that share a common synthetic strategy.
- *Fill-a-Plate* extends Route Popularity with a route diversity filter [Blaschke et al., 2020] that memorizes route frequencies across batches. Once a predefined "plate" capacity is reached, additional molecules following that route are penalized, encouraging exploration of alternative synthetic strategies.

All multi-parameter optimization (MPO) RL experiments were conducted in triplicate using REINVENT with the ChEMBL-trained prior [Loeffler et al., 2024, Mendez et al., 2019]. Training was run with a batch size of 128 for 1000 epochs.

For SFScore, reaction space was constrained to 3 NameRxn reaction classes: Suzuki coupling reactions (3.1), amide formation reactions (2.1), and N-arylations (1.3), with a maximum of 3 synthetic steps per route. Feasible hits were defined as molecules synthesizable from those reaction classes using Enamine Building Blocks as starting materials and exceeding ROCS > 0.6 (relative to the native COX-2 inhibitor SC-558 [Kurumbail et al., 1996, Dodds et al., 2024]) and QED > 0.7 [Bickerton et al., 2012], establishing a pharmaceutically relevant benchmark for synthesis-aware molecular design.

For RRScore, we used a fixed two-step reference route (2.1.10 Carboxylic ester + amine reaction, 3.1.2 Chloro Suzuki coupling). Molecules were classified as reference route hits if they could be synthesized via this route from Enamine Building Blocks, while also exceeding QED > 0.7 and ROCS > 0.6.

For Route Popularity, we quantified its effect on route coherence within batches. Route entropy was calculated as the Shannon entropy of the distribution of route signatures within each batch, providing a measure of route diversity explored by the model.

In case of Fill-a-Plate, plate capacity was set to 1000, and hits were defined as molecules synthesizable in maximum 3 steps from Enamine Building Blocks while also exceeding QED > 0.7 and ROCS > 0.6.

In all experimental runs, RL optimization was performed against a geometric mean reward comprising SynthSense, QED, and ROCS, each with equal weight. In the control runs, the weight of SynthSense was set to 0. Thus, SynthSense rewards were still evaluated to log synthetic metrics for comparison but didn't contribute to the MPO score.

3 Results and Discussion

SFScore strongly enriched for synthetically feasible molecules relative to the control baseline. Across triplicate runs, SFScore generated $6{,}237 \pm 572$ feasible

hits compared to 1,000 ± 290 for the control, representing a ∼6.2-fold increase. Feasible hit accumulation analysis showed sustained linear growth over epochs without saturation. Scaffold-level evaluation confirmed that SFScore also expanded feasible chemical space, producing 449 ± 31 unique feasible scaffolds compared to 218 ± 18 for the control (∼2.1-fold improvement).

RRScore optimization successfully directed generation toward a two-step user-defined pathway (2.1.10 carboxylic ester + amine reaction, 3.1.2 chloro Suzuki coupling). The average TED between generated routes and the reference dropped from 8.5 ± 1.3 at the start to 0.05 ± 0.06 in the final epoch, whereas the control TED even slightly increased to 9.7 ± 0.8. RRScore generated 727 ± 179 reference route hits, compared to just 1 hit in the control (∼727-fold enrichment). Structural diversity remained high, with 22.3 ± 2.5 unique scaffolds identified among reference route hits, compared to 1 in the control (∼22-fold improvement).

At the batch level, intrinsic rewards shaped the synthetic distribution of generated molecules. Route Popularity reduced route entropy within each batch, concentrating design on fewer, productive synthetic strategies. Entropy values decreased from 7.00 ± 0.11 bits at the start to 3.00 ± 0.41 bits in the final epoch, nearly two-fold lower than the control (5.75 ± 0.55 bits), indicating faster convergence on coherent synthetic solutions. In contrast, Fill-a-Plate promoted exploration by expanding the number of distinct synthetic strategies pursued. It filled 26.7 ± 3.1 plates per run compared to 11.7 ± 2.1 for the control (∼2.3-fold improvement), while maintaining a comparable hit density per plate (median 58 vs. 56). Although plates contained fewer scaffolds on average under Fill-a-Plate (14 vs. 23), the much broader synthetic coverage resulted in greater overall scaffold diversity.

Taken together, extrinsic rewards (SFScore, RRScore) enriched the synthetic feasibility of individual molecules and their alignment with target synthetic routes, while intrinsic rewards (Route Popularity, Fill-a-Plate) promoted route coherence and diversification. Across all rewards, SynthSense preserved drug-likeness and ligand similarity while substantially increasing the proportion of molecules meeting practical synthetic criteria, expanding generative design from abstract chemical space into actionable synthetic space.

In conclusion, SynthSense offers a flexible way for embedding retrosynthetic knowledge into generative molecular design. By combining intrinsic and extrinsic RL rewards, it transforms synthesizability into a controllable design parameter rather than a *post hoc* filter. The demonstrated gains in feasible hit rates, route coherence, and parallel synthesis potential highlight its practical utility for real-world discovery workflows. Future extensions could link SynthSense with automated synthesis or reaction condition prediction tools, paving the way for closed-loop design–make–test–analyze pipelines driven by generative AI.

References

G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, and A. L. Hopkins. Quantifying the chemical beauty of drugs. *Nat. Chem.*, 4:90–98, 2012.

T. Blaschke, J. Arús-Pous, H. Chen, C. Margreitter, C. Tyr-chan, O. Engkvist, K. Papadopoulos, and A. Patronov. Reinvent 2.0 – an ai tool for de novo drug design. *J. Chem. Inf. Model.*, 60:5918–592, 2020.

C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen. Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.*, 3(5):434–443, 2017.

C. W. Coley, L. Rogers, W. H. Green, and K. F. Jensen. Scscore: Synthetic complexity learned from a reaction corpus. *J. Chem. Inf. Model.*, 58(2):252–261, 2018.

M. Cretu, C. Harris, I. Igashov, A. Schneuing, M. Segler, B. Correia, J. Roy, E. Bengio, and P. Liò. Synflownet: Design of diverse and novel molecules with synthesis constraints, ver. 1. arXiv, May 2, 2024, 2024.

M. Dodds, J. Guo, T. Löhr, A. Tibo, O. Engkvist, and J. P. Janet. Sample efficient reinforcement learning with active learning for molecular design. *Chem. Sci.*, 15:4146–4160, 2024.

P. Ertl and A. Schuffenhauer. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.*, 1:8, 2009.

R. G. Kurumbail, A. M. Stevens, J. K. Gierse, J. J. McDonald, R. A. Stegeman, J. Y. Pak, D. Gildehaus, J. M. Miyashiro, T. D. Penning, K. Seibert, P. C. Isakson, and W. C. Stallings. Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents. *Nature*, 384:644–648, 1996.

H. H. Loeffler, J. He, A. Tibo, J. P. Janet, A. Voronov, L. H. Mervin, and O. Engkvist. Reinvent 4: Modern ai–driven generative molecule design. *J. Cheminf.*, 16(1):20, 2024.

D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, and A. R. Leach. Chembl: Towards direct deposition of bioassay data. *Nucleic Acids Res.*, 47: D930–D940, 2019.

NextMove. Namerxn (expert system for named reaction identification and classification). `https://www.nextmovesoftware.com/namerxn`, 2022.

M. Oestreich, I. Ewert, and M. Becker. Small molecule autoencoders: architecture engineering to optimize latent space utility and sustainability. *J. Cheminf.*, 16(1):26, 2024.

M. Oestreich, E. Merdivan, M. Lee, J. L. Schultze, M. Piraud, and M. Becker. Drugdiff: small molecule diffusion model with flexible guidance towards molecular properties. *J. Cheminf.*, 17(1):23, 2025.

M. Olivecrona, T. Blaschke, O. Engkvist, and H. Chen. Molecular de-novo design through deep reinforcement learning. *J. Cheminf.*, 9:48, 2017.

M. Pawlik and N. Augsten. Efficient computation of the tree edit distance. *ACM Trans. Database Syst.*, 40(1):3, 2015.

M. Pawlik and N. Augsten. Tree edit distance: Robust and memory-efficient. *Inf. Syst.*, 56:157–173, 2016.

P. G. Polishchuk, T. I. Madzhidov, and A. Varnek. Estimation of the size of drug-like chemical space based on gdb-17 data. *J. Comput. Aided Mol. Des.*, 27(8):675–679, 2013.

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3-4):229–256, 1992. doi: 10.1007/BF00992696.