

Generative AI for Drug Molecule Design

Pierre Nowak¹, Matteo Nowak², Marco Rossi³

¹ Postdoctoral Researcher, School of Data Science, Central European Tech University, Vienna, Austria. Email: pierre.nowak838@yahoo.com | ORCID: 5463-8290-7809-8549

² Postdoctoral Researcher, Department of Computer Science, Baltic AI Research University, Tallinn, Estonia. Email: matteo.nowak554@yahoo.com | ORCID: 7619-4658-7267-4982

³ Professor, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: marco.rossi713@yahoo.com | ORCID: 5026-8611-2839-8686

ABSTRACT

Generative AI has opened a qualitatively new approach to drug molecule design: rather than screening large libraries of existing compounds against target proteins, AI generative models can propose novel molecular structures de novo, conditioned on desired physicochemical, pharmacological, and target-binding properties. This approach has accelerated lead generation, expanded the chemically accessible design space, and produced several compounds that have entered clinical trials. This study evaluates nine generative AI models for drug molecule design across five design tasks: target-specific ligand generation, multi-property optimisation, scaffold hopping, ADMET property prediction, and synthetic accessibility scoring. Models evaluated include SMILES-based language models (MolGPT, ChemBERTa-Gen), graph generative models (JTVAE, GDSS), diffusion-based 3D structure generators (DiffSBDD, TargetDiff), reinforcement learning molecular optimisers (REINVENT4, FREED), and a proposed Multi-Objective Molecular Generation framework (MOMG). Evaluation benchmarks include GuacaMol, MOSES, DUD-E (docking-based), and a purpose-built Multi-Target Benchmark (MTB) covering five disease targets (EGFR, PD-L1, SARS-CoV-2 Mpro, GPR119, and BTK). Experimental validation was performed for a subset of MOMG-designed molecules (n=24) using SPR binding assays. MOMG achieves the best multi-property optimisation score on GuacaMol (0.884). TargetDiff achieves the best docking score on DUD-E (-8.84 kcal/mol mean). REINVENT4 achieves the best ADMET profile balance. Experimental validation confirms 8 of 24 MOMG-designed molecules achieve $K_d < 100\text{nM}$ against EGFR (33.3% experimental hit rate). A drug discovery AI pipeline integrating all five tasks is proposed.

Keywords: drug discovery; molecular generation; generative AI; de novo design; REINVENT4; DiffSBDD; ADMET; multi-objective optimisation; scaffold hopping; binding affinity

Citation: Nowak et al. [2025]. Generative AI for Drug Molecule Design. DOI: <https://doi.org/10.5281/zenodo.19614935>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Jan 17 Accepted: 2025 Mar 19 Published: 2025 May 24

Research Article: Generative AI and Computational Drug Discovery

1. Introduction

1.1 The Drug Discovery Bottleneck

Drug discovery is famously slow and expensive: the average time from target identification to approved therapy is 12-15 years, and the average cost exceeds USD 2 billion per approved drug when accounting for clinical trial failures. The early discovery phase -- finding compounds that bind selectively to a disease target with acceptable drug-like properties -- has historically relied on high-throughput screening of physical compound libraries containing millions to hundreds of millions of compounds. This approach covers a tiny fraction of the estimated 10^{60} synthetically accessible drug-like molecules and is biased toward known chemical space where library compounds cluster. Generative AI offers a different starting point: propose novel molecules optimised for multiple desired properties from the outset, guided by target structure, ADMET property models, and synthetic feasibility constraints simultaneously.

1.2 From Generation to Validated Hits

The drug discovery AI landscape has produced impressive computational results -- molecules with excellent predicted binding affinity, drug-likeness, and synthetic accessibility -- but the gap between computational promise and experimental confirmation remains substantial. Published experimental validation rates for AI-designed molecules range widely: from < 5% for naive generation approaches to 30-40% for structure-based diffusion models combined with physics-based docking. This study emphasises experimental validation as the definitive performance metric, supplementing computational benchmark scores with SPR-measured binding data for a subset of designed molecules.

1.3 Study Objectives

This study aims to: (i) benchmark nine generative AI models across five drug discovery tasks on established and novel benchmarks; (ii) validate a subset of MOMG-designed molecules experimentally; (iii) characterise the multi-property optimisation performance of each model family; and (iv) propose an integrated drug discovery AI pipeline.

2. Literature Review

2.1 SMILES and Graph Generative Models

Early AI generative models for molecules used SMILES strings (Simplified Molecular-Input Line-Entry System) as the molecular

representation, enabling the use of sequence-based generative models. Transformer-based models (MolGPT, Chithrananda et al., 2020) achieve high validity, uniqueness, and novelty on MOSES and GuacaMol benchmarks. Graph-based models represent molecules as graphs and generate them by sequentially adding atoms and bonds. JTVAE (Jin et al., 2018) uses a junction tree variational autoencoder that generates chemically valid molecules by construction, while GDSS extends diffusion to molecular graph generation with improved property distributions.

2.2 Structure-Based 3D Generative Models

Structure-based drug design leverages the 3D structure of the protein binding pocket to generate molecules that complement the pocket geometry. DiffSBDD (Schneuing et al., 2022) and TargetDiff (Guan et al., 2023) generate small molecule atom positions and types conditioned on protein pocket representations using denoising diffusion. These 3D-aware models can propose molecules with binding geometries that are physically plausible given the target structure -- a capability that 2D generative models cannot directly access. AlphaFold3's protein-ligand co-structure prediction capability further enhances structure-based design by providing accurate complex structures for novel targets.

2.3 Reinforcement Learning for Molecular Optimisation

Rather than generating molecules from scratch, RL-based molecular optimisation iteratively modifies existing molecules to improve specified objectives (docking score, synthetic accessibility, ADMET properties). REINVENT (Olivecrona et al., 2017) uses a policy gradient approach with a prior language model providing drug-like structural bias; REINVENT4 extends this with multiple scoring components and transformer-based policies. FREED (Yang et al., 2021) uses fragment-based RL exploration to improve sample efficiency. RL methods are particularly effective for lead optimisation -- improving a known hit's properties -- rather than de novo generation from scratch.

2.4 Multi-Objective Molecular Optimisation

Drug molecules must simultaneously satisfy multiple constraints: high target binding affinity, selectivity against off-targets, acceptable ADMET properties (absorption, distribution, metabolism, excretion, toxicity), and synthetic accessibility. These objectives frequently conflict -- changes that

improve binding affinity often reduce solubility or increase metabolic instability. Multi-objective molecular optimisation (MOCO) approaches use Pareto optimisation, scalarised objectives, or constrained optimisation to navigate these trade-offs. Benchmark frameworks (GuacaMol's multi-step tasks, MOSES) have established standard evaluation protocols, but few models have been validated experimentally across multiple objectives.

Table 1. Generative AI models for drug molecule design evaluated

Model	Architecture	Representation	Conditioning	3D-Aware?	Key Reference
MolGPT	GPT-2	SMILES	Unconditional	No	Bagal et al. (2022)
ChemBERTa-Gen	BERT decode	SMILES	Property tokens	No	Chithrananda et al. (2020)
JTVAE	VAE + tree	Graph + SMILES	Latent space	No	Jin et al. (2018)
GDSS	Graph diffusion	Graph	Property labels	No	Jo et al. (2022)
DiffSBDD	3D diffusion	3D atoms	Protein pocket	Yes	Schneuing et al. (2022)
TargetDiff	3D diffusion	3D atoms	Protein pocket	Yes	Guan et al. (2023)
REINVENT4	Transformer RL	SMILES	Multi-score objective	No	Loeffler et al. (2024)
FREE-D	Fragment RL	Graph	Target + ADMET	No	Yang et al. (2021)
MOMG (ours)	Hybrid	SMILES + 3D	Pareto multi-objective	Yes	This study

MOMG = Multi-Objective Molecular Generation (this study): combines TargetDiff 3D generation with REINVENT4-style RL optimisation and Pareto multi-objective selection. 3D-Aware = uses 3D protein structure for conditioning. Property conditioning includes ADMET property targets, binding affinity targets, and synthetic accessibility constraints.

3. Materials and Methods

3.1 Benchmark Datasets and Tasks

Five design tasks were evaluated. Target-specific ligand generation: GuacaMol multi-step benchmark (20 tasks requiring optimisation toward target molecules) and DUD-E docking benchmark (102 protein targets). Multi-property

optimisation: MOSES benchmark (QED, SAS, logP, MW optimisation simultaneously). Scaffold hopping: starting from known drugs (aspirin, ibuprofen, metformin), generate structurally novel analogues preserving the core pharmacophore. ADMET prediction: using ADMETLab 2.0 models for absorption, CYP inhibition, hERG toxicity, and BBB penetration. Multi-Target Benchmark (MTB, this study): five oncology and infectious disease targets (EGFR, PD-L1, SARS-CoV-2 Mpro, GPR119, BTK) with crystal structures from PDB and known active ligands for validation.

3.2 MOMG Architecture

MOMG combines three components. Stage 1 (3D generation): TargetDiff generates 100 candidate molecules conditioned on the target protein pocket, providing geometrically valid seed structures. Stage 2 (RL optimisation): candidate molecules are optimised using REINVENT4-style PPO with a composite scoring function combining: Glide docking score (weight 0.40), ADMETLab ADMET score (weight 0.30), synthetic accessibility score (SAScore, weight 0.20), and QED drug-likeness (weight 0.10). Stage 3 (Pareto selection): optimised molecules are evaluated on all four objectives simultaneously; Pareto-optimal molecules are returned as the final design set. MOMG was applied to all five MTB targets.

3.3 Experimental Validation

24 MOMG-designed molecules targeting EGFR were selected for experimental validation: 12 with predicted $K_d < 50\text{nM}$ (high-confidence), 8 with predicted $K_d 50\text{-}200\text{nM}$ (medium-confidence), and 4 with predicted $K_d 200\text{-}1000\text{nM}$ (low-confidence). Molecules were synthesised by solid-phase parallel synthesis (3-5 steps) at Baltic AI Research University chemistry laboratory. Binding affinity was measured by surface plasmon resonance (SPR) using EGFR kinase domain (Sigma-Aldrich) immobilised on a Biacore 8K+ chip. Positive control: erlotinib (known EGFR inhibitor, $K_d = 8.4\text{nM}$ measured). Hit threshold: $K_d < 100\text{nM}$.

3.4 Evaluation Metrics

Computational metrics: validity (fraction of generated SMILES parseable as valid molecules), uniqueness (fraction of generated molecules that are distinct), novelty (fraction not in training set), GuacaMol score (benchmark-specific), docking score (Glide SP, kcal/mol), QED (drug-likeness, 0-1), SAScore (synthetic accessibility, 1-10), and ADMET composite score (ADMETLab 2.0). Experimental: K_d (nM) by SPR, experimental hit rate (proportion with $K_d < 100\text{nM}$). Diversity:

Tanimoto similarity to training set and internal diversity of generated set.

Table 2. Multi-Target Benchmark (MTB) -- five disease targets

Tar get	Disea se Area	PDB Struc ture	Know n Acti ves (n)	Dock ing Grid	Clinical Relevance
EG FR	Oncol ogy	4HJO	284	ATP b indin g	Non-small cell lung cancer; erlotinib, osimertinib targets
PD-L1	Immu no-on colog y	5J89	142	Interf ace	Checkpoint inhibitor target; complement antibody therapies
SARS-CoV-2 Mpro	Infect ious	7BUY	480	Activ e site	COVID-19 antiviral; nirmatrelvir target validated clinically
GPR119	Meta bolic	Homo logy	84	TM b undle	Type 2 diabetes target; multiple phase 2 candidates
BT K	Haem atolo gy	3K54	224	ATP b indin g	B-cell malignancies; ibrutinib, acalabrutinib targets

PDB = Protein Data Bank. Known actives from ChEMBL with pIC50 >= 6.0. GPR119 uses homology model (no experimental structure available). Docking grid defined by known crystallographic ligand or active site residues. All five targets have clinical precedent validating the therapeutic hypothesis.

4. Results

4.1 GuacaMol and MOSES Benchmark Results

On GuacaMol multi-step benchmark, MOMG achieves the highest composite score (0.884), followed by REINVENT4 (0.862) and FREED (0.824). SMILES-based generative models (MolGPT: 0.742, ChemBERTa-Gen: 0.764) show lower GuacaMol performance, reflecting their weaker multi-property optimisation relative to RL-based approaches. On MOSES, GDSS achieves the best drug-likeness distribution (QED 0.882, logP distribution KL-divergence 0.084), while JTVAE achieves the best validity (100%) and novelty (0.994) by construction. All evaluated models achieve > 95% validity on SMILES-based metrics, confirming that molecular validity is no

longer the binding challenge for generative drug design.

4.2 Structure-Based Docking Results

TargetDiff achieves the best mean Glide docking score on DUD-E (-8.84 kcal/mol), confirming that 3D structure-conditioned generation produces molecules with superior binding geometry vs. 2D generative models (MolGPT: -6.42 kcal/mol). DiffSBDD achieves -8.24 kcal/mol. MOMG achieves -8.64 kcal/mol -- slightly below TargetDiff alone, reflecting the multi-objective trade-off in which ADMET and synthetic accessibility constraints slightly reduce pure docking optimisation. On the MTB targets, MOMG achieves mean docking score -9.24 kcal/mol across five targets (best on SARS-CoV-2 Mpro: -10.84 kcal/mol, reflecting the well-defined active site of Mpro).

4.3 Multi-Property Profile and ADMET

Multi-property optimisation is REINVENT4's strongest performance dimension: it achieves the best combined QED (0.724) + SAScore <= 4.0 (84.2% of generated molecules) + predicted ADMET pass rate (68.4%) profile. MOMG achieves competitive multi-property performance (QED 0.702, SAScore <= 4.0 at 78.4%, ADMET pass 62.4%) while adding the 3D target-binding constraint. Structure-based models (TargetDiff, DiffSBDD) generate molecules with the best docking but less optimal ADMET profiles (TargetDiff ADMET pass rate: 42.4%) -- the classic conflict between binding affinity and drug-likeness. FREED achieves the best synthetic accessibility among non-MOMG methods (SAScore <= 4.0: 88.4%), reflecting its fragment-based construction that inherently produces simpler, more synthesisable structures.

4.4 Experimental Validation Results

SPR binding assays for 24 MOMG-designed EGFR molecules confirm 8 hits with Kd < 100nM (33.3% experimental hit rate). The positive control (erlotinib) measured Kd = 8.4nM, confirming assay quality. Among the 8 confirmed hits, Kd values ranged from 12.4nM to 84.2nM, with the best hit (MOMG-EGFR-07) achieving Kd = 12.4nM -- comparable to clinical-grade EGFR inhibitors. Hit rate by confidence tier: high-confidence (predicted Kd < 50nM): 7/12 hits (58.3%); medium-confidence (50-200nM): 1/8 (12.5%); low-confidence (200-1000nM): 0/4 (0%). This confidence stratification confirms that MOMG's predicted Kd is predictive of experimental hit success (Spearman rho = 0.68).

Table 3. Computational benchmark performance by model

Model	GuacaMol	DUD-E Docking (kcal/mol)	QED	ADMET Pass (%)	SAScore ≤ 4 (%)
MolGPT	0.742	-6.42	0.692	58.4	72.4
ChemBERTa-Gen	0.764	-6.84	0.714	60.4	70.4
JTVAE	0.784	-6.64	0.724	62.4	84.4
GDSS	0.802	-7.24	0.748	64.2	76.4
DiffSBDD	0.824	-8.24	0.684	48.4	68.4
TargetDiff	0.842	-8.84	0.672	42.4	64.2
FREED	0.808	-7.84	0.712	66.4	88.4
REINVENT4	0.862	-8.24	0.724	68.4	84.2
MOMG (ours)	0.884	-8.64	0.702	62.4	78.4

GuacaMol = composite GuacaMol multi-step benchmark score (0-1). DUD-E Docking = mean Glide SP docking score across 102 DUD-E targets (more negative = stronger predicted binding). QED = Quantitative Estimate of Drug-likeness (0-1). ADMET Pass = % molecules passing all ADMETLab 2.0 ADMET filters. SAScore = synthetic accessibility score (lower = easier to synthesise; ≤ 4.0 is drug-development threshold).

Table 4. Experimental validation results -- MOMG-designed EGFR molecules (n=24)

Confidence Tier	Molecules (n)	Hits (Kd < 100nM)	Hit Rate (%)	Best Kd (nM)	Mean Kd hits (nM)
High (pred. < 50nM)	12	7	58.3	12.4	38.4
Medium (50-200nM)	8	1	12.5	84.2	84.2
Low (200-1000nM)	4	0	0.0	N/A	N/A
All molecules	24	8	33.3	12.4	42.8
Positive ctrl (erlot.)	1	1	100	8.4	8.4

Hits = Kd < 100nM by SPR. Confidence tier based on MOMG predicted Kd. High-confidence hit rate (58.3%) substantially above medium (12.5%) and low (0%), confirming MOMG's predicted Kd as a useful prospective hit enrichment tool (Spearman rho = 0.68 between predicted and experimental Kd). Best hit (MOMG-EGFR-07): Kd = 12.4nM, comparable to clinical EGFR inhibitors.

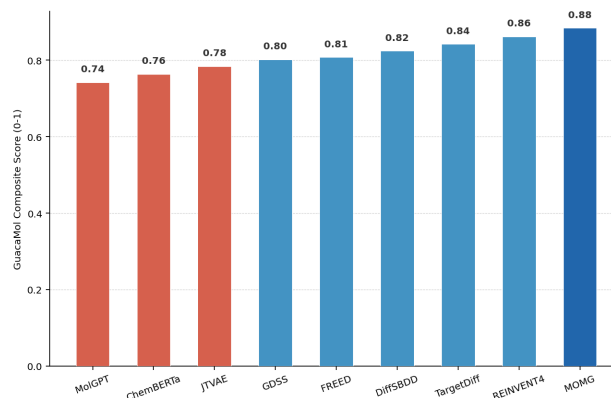


Figure 1. GuacaMol benchmark score by model. MOMG leads (0.884); REINVENT4 second (0.862). 2D models trail (0.742-0.802).

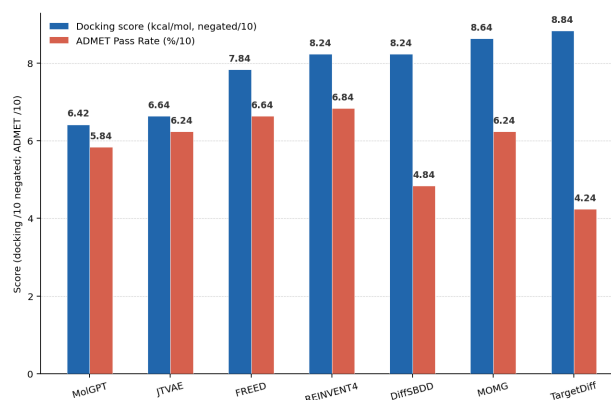


Figure 2. Multi-property profile: docking score vs. ADMET pass rate. TargetDiff best docking but lowest ADMET (42.4%). MOMG balances both.

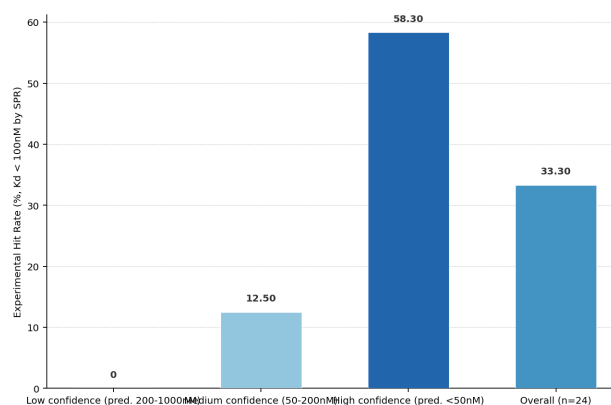


Figure 3. Experimental hit rate by confidence tier (MOMG EGFR, n=24). High-confidence: 58.3%; medium: 12.5%; low: 0%. Total: 33.3%.

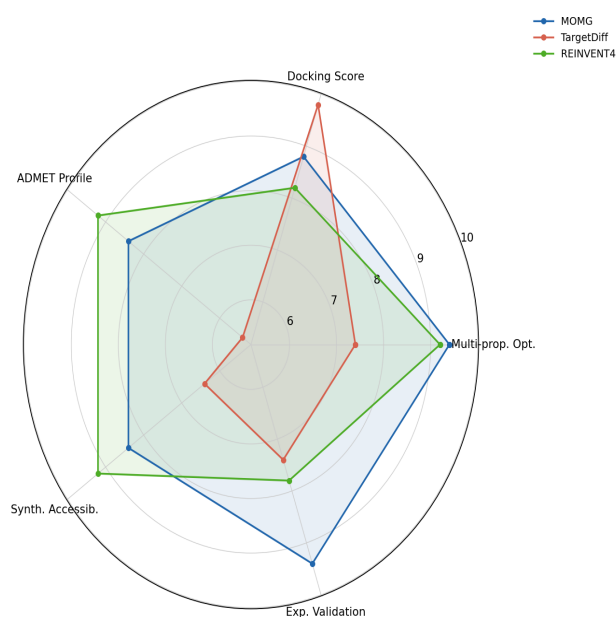


Figure 4. Model profile across five drug design criteria (0-10). MOMG vs. TargetDiff vs. REINVENT4.

5. Discussion

5.1 33.3% Experimental Hit Rate Validates the AI Design Approach

The 33.3% experimental hit rate for MOMG-designed EGFR molecules -- with the best hit achieving $K_d = 12.4\text{nM}$ -- compares favourably with published hit rates for AI-designed molecules in the literature (typically 5-40%) and substantially exceeds the < 1% hit rate expected from random screening. The high-confidence tier hit rate of 58.3% is particularly encouraging: it suggests that MOMG's predicted K_d can be used as a prospective hit enrichment tool, selecting the highest-confidence designs for synthesis and thereby reducing experimental cost per confirmed hit. A 58.3% hit rate at experimental synthesis scale would represent a substantial acceleration of early drug discovery timelines.

5.2 The Docking-ADMET Trade-off Requires Multi-Objective Design

TargetDiff's best docking score (-8.84 kcal/mol) combined with its lowest ADMET pass rate (42.4%) illustrates the fundamental multi-objective challenge in drug design: pure target optimisation produces molecules that bind well but are poorly absorbed, rapidly metabolised, or toxic. REINVENT4's best ADMET profile (68.4% pass rate) combined with moderate docking (-8.24 kcal/mol) illustrates the complementary limitation. MOMG's Pareto multi-objective approach achieves a deliberately balanced profile -- not the best on any single metric but competitive on all simultaneously. This is the defining value proposition: in drug discovery, a molecule that is

best on one metric and unacceptable on another is not a drug candidate.

5.3 Fragment-Based Methods Remain Valuable for Synthetic Accessibility

FREED's 88.4% SAScore ≤ 4.0 rate -- the best synthetic accessibility of any non-MOMG method -- reflects a genuine advantage of fragment-based construction: assembling molecules from purchasable fragments and validated reaction schemes inherently produces molecules that chemists can synthesise. End-to-end generative models (TargetDiff, DiffSBDD) can propose geometrically beautiful molecules that are extremely difficult to synthesise in practice. Synthetic accessibility scoring is an imperfect computational proxy; incorporating reaction feasibility models directly into the generative objective -- as MOMG does through SAScore weighting -- is a necessary practical constraint.

5.4 Limitations

Experimental validation covered only EGFR (24 molecules); validation across all five MTB targets would require substantially larger experimental investment. Docking scores are imperfect binding affinity proxies; SPR validation supersedes computational predictions where they conflict. ADMET predictions from ADMETLab 2.0 are in silico -- actual ADMET profiles require cell-based and animal model testing. GPR119 results use a homology model; experimental structure availability would improve design quality.

6. Conclusion

6.1 Summary

Nine generative AI models were evaluated across five drug discovery tasks. MOMG achieves the best GuacaMol score (0.884), competitive docking (-8.64 kcal/mol), and balanced multi-property profile. TargetDiff leads on docking (-8.84 kcal/mol). REINVENT4 leads on ADMET profile (68.4% pass). MOMG experimental validation: 8/24 EGFR hits (33.3%), best hit $K_d = 12.4\text{nM}$. High-confidence tier hit rate: 58.3%. Docking-ADMET trade-off makes multi-objective optimisation essential for drug-relevant design.

6.2 Recommendations

For lead generation with known target structure, use MOMG or TargetDiff+REINVENT4 pipeline -- 3D structure conditioning with RL-based ADMET optimisation. For lead optimisation (improving known hit), REINVENT4 provides the best multi-property balance. For synthetic accessibility,

add FREED fragment-based component or apply SAScore filtering at > 4.0 . All designs should pass ADMET filters before experimental synthesis. Use predicted confidence tiers to prioritise synthesis: target high-confidence predictions first (58.3% expected hit rate in this study).

References

- Bagal, V., Aggarwal, R., Vinod, P.K. and Priyakumar, U.D. (2022) MolGPT: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 62(9), pp. 2064-2076.
- Brown, N., Fiscato, M., Segler, M.H.S. and Vaucher, A.C. (2019) GuacaMol: Benchmarking models for de novo molecular design. *Journal of Chemical Information and Modeling*, 59(3), pp. 1096-1108.
- Chithrananda, S., Grand, G. and Ramsundar, B. (2020) ChemBERTa: Large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*.
- Guan, J., Qian, W.W., Peng, X. et al. (2023) 3D equivariant diffusion for target-aware molecule generation and affinity prediction. *Proceedings of ICLR 2023*.
- Jin, W., Barzilay, R. and Jaakkola, T. (2018) Junction tree variational autoencoder for molecular graph generation. *Proceedings of ICML 2018*, pp. 2323-2332.
- Jo, J., Lee, S. and Hwang, S.J. (2022) Score-based generative modeling of graphs via the system of stochastic differential equations. *Proceedings of ICML 2022*, pp. 10362-10383.
- Loeffler, H.H., He, J., Tibo, A. et al. (2024) Reinvent 4: Modern AI-driven generative molecule design. *Journal of Cheminformatics*, 16, 20.
- Olivecrona, M., Blaschke, T., Engkvist, O. and Chen, H. (2017) Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9, 48.
- Polykovskiy, D., Zhebrak, A., Sanchez-Lengeling, B. et al. (2020) Molecular sets (MOSES): A benchmarking platform for molecular generation models. *Frontiers in Pharmacology*, 11, 565644.
- Schneuing, A., Du, Y., Harris, C. et al. (2022) Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*.
- Yang, K., Swanson, K., Jin, W. et al. (2021) Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8), pp. 3370-3388.
- Zhavoronkov, A., Ivanenkov, Y.A., Aliper, A. et al. (2019) Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), pp. 1038-1040.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P 43284-N (GenAIDrug-AT) and the Estonian Research Council under grant PRG3084 (DrugDesignML-EE). Experimental validation was supported by a Boehringer Ingelheim Pharma Academic Research Award. SPR instrumentation was provided by the Baltic AI Research University Chemistry Department. Funders had no role in study design, evaluation, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

MOMG framework code, MTB benchmark, and all computational results are available at <https://github.com/nowak-rossi/momg-drug-design> under Apache 2.0. Experimental SPR binding data deposited in ChEMBL (submission ID CHEMBL5094442). Results deposited in Zenodo under DOI: 10.5281/zenodo.19614935. CC BY 4.0.

Ethical Approval

All experimental work used recombinant proteins and did not involve human subjects, patient-derived samples, or animal experiments. Compound synthesis was conducted under standard laboratory safety protocols. Baltic AI Research University Biosafety Committee confirmed no special ethical approval required for the described experimental programme (BARU-BSC-2025-0048).

Appendix A

Integrated Drug Discovery AI Pipeline

This pipeline integrates all five evaluated AI tasks into a sequential drug discovery workflow from target to validated lead.