Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

MOLECULAR FOUNDATION MODELS FOR LEAD OPTIMIZATION USING BIOACTIVITY, ADMET, AND SYNTHETIC FEASIBILITY PROMPTS

Peter Novak[1]*, Jana Svoboda[1]

1. *Department of Drug Discovery Informatics, Faculty of Pharmacy, Charles University, Prague, Czech Republic.*

ARTICLE INFO

Received:
08 November 2025
Received in revised form:
28 January 2026
Accepted:
29 January 2026
Available online:
28 February 2026

Keywords: Molecular foundation models, Generative chemistry, Prompt-conditioned generation, Lead optimization, ADMET, Synthetic feasibility

ABSTRACT

Lead optimization involves the simultaneous enhancement of potency, ADMET properties, and synthetic feasibility, making the progression from an initial hit or lead to a viable drug candidate a challenging multi-objective design problem. Traditional medicinal chemistry workflows remain iterative, expert-intensive, and reliant on repeated cycles of design, synthesis, and testing, while existing molecular generative models often target single-property optimization or employ reward functions that, though powerful, are not always intuitive for medicinal chemists to guide. To address these limitations, this article proposes a molecular foundation model for prompt-conditioned lead optimization, designed to generate optimized lead molecules from natural-language or structured prompts specifying desired bioactivity, ADMET, and synthetic feasibility constraints. The system leverages a pre-trained transformer-based molecular language model fine-tuned for conditional generation, where a prompt encoder directs molecule generation toward the requested target profile, and reinforcement learning aligns outputs with bioactivity, ADMET, and synthesis-oriented reward signals. The model aims to produce a small, diverse set of chemically valid candidates tailored to the prompt rather than an exhaustive random library, providing medicinal chemists with a curated selection for review. By combining chemical language modeling with multi-objective reward design, prompt-conditioned molecular foundation models have the potential to make lead optimization more interactive, transparent, and parallelizable, supporting more efficient exploration of drug-like chemical space.

To Cite This Article: Novak P, Svoboda J. Molecular Foundation Models for Lead Optimization Using Bioactivity, ADMET, and Synthetic Feasibility Prompts. Pharmacophore. 2026;17(1):53-61. https://doi.org/10.51847/TzZy4qWv97

Introduction

Lead optimization is a multi-objective process in which potency must be improved without compromising physicochemical properties, safety liabilities, or synthetic tractability. Reinforcement-learning frameworks for de novo design showed that molecular generation can be guided toward desired target profiles [1], while later multi-parameter systems emphasized that medicinal chemistry optimization is rarely reducible to a single property [2]. In practical drug discovery, each proposed analog must also pass through design, synthesis, purification, testing, and interpretation before the next design cycle begins. This iterative structure motivates computational systems that can propose more focused analog sets before laboratory resources are committed.

Generative chemistry has expanded from recurrent neural networks and variational autoencoders to graph models, transformers, and reinforcement-learning agents. Early recurrent approaches demonstrated focused molecular library generation from chemical language models [3], while continuous latent representations enabled interpolation and optimization in molecular space [4]. Conditional graph generation further showed that molecular design could be shaped by specified objectives [5], but many systems still risk emphasizing easily optimized scores over synthesis, assay relevance, or medicinal chemistry interpretability. Benchmarking efforts such as GuacaMol [6] and MOSES [7] helped clarify that validity and novelty alone are insufficient for lead optimization.

Foundation models introduce a different framing: a molecular model is first pre-trained to capture broad chemical syntax and structure-property regularities, then adapted to downstream design tasks. MolGPT illustrated how transformer decoders can generate molecules in a language-modeling framework [8], while molecule-text systems showed that natural language can be

Corresponding Author: Peter Novak; Department of Drug Discovery Informatics, Faculty of Pharmacy, Charles University, Prague, Czech Republic. E-mail: peter.novak@gmail.com.

aligned with molecular representations for retrieval and editing [9]. More recent molecular foundation models and prompt-oriented systems extend this idea toward conditional generation from richer molecular or textual context [10, 11]. These developments suggest that lead optimization could be expressed as a prompt-guided generation problem rather than a manually reconfigured model-training task.

The central thesis of this EAI article is that a molecular foundation model could accept a multi-property prompt and return a focused set of candidate analogs for lead optimization. Such a system would combine a pre-trained SMILES decoder [8], prompt-conditioned generation [11], multi-constraint optimization [12], and synthesis-aware scoring [13, 14] into a single interactive interface. Instead of asking chemists to tune separate models for potency, ADMET, and synthetic feasibility, the prompt would specify the intended optimization profile. The model would then be expected to generate candidate molecules whose final selection remains under expert medicinal chemistry judgment.

Background

Principles of Lead Optimization

Lead optimization typically moves from potency improvement to broader assessment of solubility, permeability, metabolic stability, clearance, off-target safety, and synthetic practicality. Deep reinforcement-learning approaches demonstrated how generated molecules can be biased toward desired activity profiles [15], while multi-parameter reinforcement-learning design emphasized the need to optimize several medicinal chemistry objectives together [2]. In practice, these objectives may be addressed sequentially when one liability dominates or in parallel when several properties must remain within acceptable ranges. A prompt-conditioned system would formalize this medicinal chemistry funnel by allowing the user to request activity preservation, ADMET improvement, and synthetic feasibility in a single generation step.

Molecular Foundation Models

Molecular foundation models are pre-trained chemical language or graph models that learn reusable representations before task-specific adaptation. SMILES-based transformers such as MolGPT treat molecular strings as generative language [8], while later training frameworks for chemical foundation models systematize the construction of reusable molecular encoders and decoders. Graph-based foundation models such as MolE illustrate the complementary idea that molecular structure can be represented beyond linear notation [16]. These models are attractive for lead optimization because pre-training can encode chemical syntax, common substructures, and broad property regularities before prompt-specific fine-tuning.

Prompt-Conditioned Generation

Prompt-conditioned molecular generation refers to steering a generative model with explicit design instructions, such as property tokens, structured fields, natural-language goals, or learned prompt embeddings. Multi-constraint transformer systems showed how generation can be conditioned by target constraints and refined with reinforcement learning [12], while natural-language molecule models demonstrated that textual descriptions can be mapped into molecular operations [9]. SAFE-based molecular design further suggests that chemically meaningful sequence representations may improve prompt controllability [11]. In a lead optimization interface, a prompt could therefore serve as the user-facing control layer for objectives that would otherwise be encoded as hidden reward weights.

Multi-Objective Optimization in Chemistry

Multi-objective molecular optimization must balance objectives that can conflict, such as potency, polarity, permeability, metabolic stability, and synthetic simplicity. Conditional graph generation provided one route for multi-objective design [5], while policy-gradient molecular generation showed how reward signals can guide transformer outputs toward property-defined regions [17]. Multi-parametric de novo design workflows further frame lead optimization as a negotiation among competing medicinal chemistry priorities rather than a search for a single best score [18]. This makes Pareto-style exploration and adjustable weighting more appropriate than rigid single-objective optimization.

Synthetic Feasibility and Its Integration

Synthetic feasibility is essential because an attractive generated molecule has limited value if it cannot be made efficiently or reliably. SCScore introduced a learned measure of synthetic complexity from reaction data [13], while RAscore connected generative design to retrosynthetic accessibility by estimating whether a molecule is likely to be reachable through AI-driven retrosynthetic planning [14]. These synthesis-oriented metrics can be used as constraints, reward modifiers, or filtering stages during molecular generation. For prompt-conditioned lead optimization, synthetic feasibility should be treated as a first-class objective rather than as a post hoc rejection criterion.

Model System Overview

High-Level Workflow

The conceptual workflow begins when a medicinal chemist provides a prompt such as retaining a lead scaffold, adding polarity, improving aqueous solubility, maintaining acceptable synthetic accessibility, and avoiding a specified liability. A pre-trained molecular generator proposes candidate SMILES, similar in spirit to transformer-based generation frameworks [8] and conditional systems designed for multi-constraint molecular output [12]. An external predictive suite then estimates

bioactivity, ADMET behavior, and synthetic feasibility, while reinforcement learning updates generation toward prompt satisfaction [1]. The loop remains conceptual in this EAI proposal and should be understood as a model-oriented workflow rather than an experimentally validated implementation.

Figure 1 presents the proposed prompt-conditioned molecular foundation model architecture, showing how medicinal chemistry intent is translated into multi-objective molecular generation, synthesis-aware reward alignment, and curated analog selection for expert review.

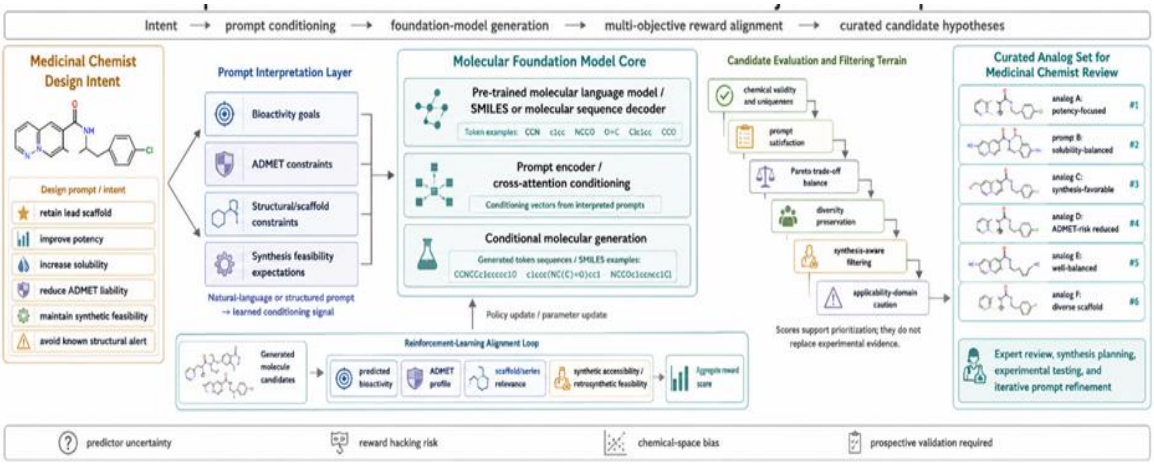

Figure 1. Prompt-Conditioned Molecular Foundation Model for Multi-Objective Lead Optimization.

Core Components

The system consists of a pre-trained SMILES decoder, a prompt encoder, property prediction modules, and a multi-objective reward calculator. The decoder provides chemical language competence as in transformer molecular generators [8], while the prompt encoder follows molecule-text alignment work showing that textual and molecular representations can be linked [9, 10]. Bioactivity and ADMET predictors supply task-specific scoring, and synthesis estimators such as SCScore [13] and retrosynthetic accessibility models [14] provide feasibility signals. The reward calculator integrates these modules into a prompt-dependent objective rather than a fixed universal scoring rule.

Table 1 decomposes the proposed system into functional layers, clarifying how prompt interpretation, conditional molecular generation, predictive scoring, synthesis-aware evaluation, and expert review interact within the conceptual lead optimization architecture.

Table 1. Architectural Logic of the Prompt-Conditioned Molecular Foundation Model for Lead Optimization

System layer	Primary function in the proposed framework	Input handled by the layer	Model or computational mechanism	Output produced	Contribution to lead optimization	Key risk if poorly designed
Medicinal chemistry prompt layer	Converts expert design intent into a machine-readable optimization request	Natural-language or structured instructions about potency, ADMET, scaffold retention, and synthesis	Prompt templates, controlled vocabulary, property tokens, or learned prompt embeddings	Explicit design profile for the generation task	Allows chemists to steer optimization without manually tuning reward equations	Ambiguous prompts may produce molecules that satisfy model-interpreted goals rather than chemist-intended goals
Prompt encoding layer	Transforms the requested target profile into conditioning information for the generator	Bioactivity goals, ADMET thresholds, structural constraints, synthesis expectations	Token embedding, learned prompt vector, prefix conditioning, or cross-attention	Prompt-conditioned latent or token-level control signal	Links human-readable medicinal chemistry priorities to molecular generation	Weak alignment between prompt semantics and molecular output may reduce controllability
Pre-trained molecular generator	Provides chemical syntax competence and broad structure-property priors	Molecular strings, SMILES-like sequences, or chemically meaningful sequence representations	Transformer-based molecular language model or related molecular foundation model	Candidate molecular sequences	Enables generation of chemically valid and drug-like analog hypotheses	Pre-training bias may favor familiar scaffolds and underrepresent novel chemical regions
Conditional generation module	Produces analogs under the constraints specified by the prompt	Prompt embedding plus molecular-generation context	Conditional decoding, scaffold-aware generation, stochastic sampling, constrained decoding	Diverse candidate analogs	Supports focused exploration of a chemical series rather than broad	Overly loose conditioning may drift away from the lead series; overly strict conditioning may

					random library generation	suppress useful novelty
Bioactivity prediction module	Estimates whether generated molecules preserve or improve target-relevant activity	Generated molecules and target-specific activity context	QSAR, docking-informed predictors, learned activity models, or project-specific models	Predicted potency or activity score	Prioritizes analogs likely to retain desired biological activity	Model artifacts may be amplified during reinforcement learning
ADMET prediction module	Estimates developability-related liabilities	Generated molecules and physicochemical descriptors	Solubility, permeability, clearance, toxicity, off-target, or alert models	ADMET profile estimate	Prevents optimization from focusing only on potency	Inaccurate ADMET predictors may reject useful molecules or retain risky ones
Synthetic feasibility module	Evaluates whether generated molecules are practical to synthesize	Generated molecules, route-likelihood estimates, synthetic complexity signals	SCScore-like complexity estimation, RAscore-like retrosynthetic accessibility, rule-based filters	Synthesis-aware feasibility score	Treats makeability as a first-class optimization objective	Proxy scores may miss reagent availability, route robustness, yield, purification burden, or project constraints
Multi-objective reward calculator	Integrates bioactivity, ADMET, scaffold relevance, diversity, and synthesis signals	Property predictions, feasibility estimates, prompt constraints	Weighted reward, Pareto-aware scoring, constraint satisfaction, penalty terms	Prompt-dependent reward signal	Balances competing medicinal chemistry objectives	Poor weighting may create reward hacking or single-property domination
Reinforcement-learning alignment module	Adjusts generation toward prompt satisfaction and reward-balanced output	Generated molecules and reward feedback	Policy-gradient or reinforcement-learning-based molecular generation	Improved conditional generation policy	Aligns candidate generation with the requested optimization profile	Reinforcement learning may exploit weak predictors rather than improving true molecule quality
Candidate curation and expert review layer	Converts generated molecules into actionable design hypotheses	Filtered and scored candidate molecules	Ranking, clustering, diversity selection, medicinal chemist inspection	Small curated analog set	Keeps human judgment central before synthesis and testing	Overtrust in model-ranked candidates may reduce critical medicinal chemistry evaluation

Design Principles

The first design principle is prompt-driven control, meaning that a chemist can modify the desired optimization profile without retraining a separate model for every target combination. The second principle is flexibility, drawing on multi-constraint generation [12], conditional molecular design [19], and dynamic multi-conditional transformer concepts [20] to support different mixtures of activity, ADMET, and synthesis goals. The third principle is human-centric deployment, because generated molecules should be treated as design hypotheses for expert review rather than automated decisions. In this sense, the model supports medicinal chemistry judgment instead of replacing it.

Prompt Engineering for Multi-Objective Lead Optimization

Prompt Structure and Semantics

A lead optimization prompt can encode the target activity profile, required ADMET filters, structural constraints, and synthetic feasibility expectations in either natural language or structured key-value form. Natural-language molecule systems demonstrate that textual descriptions can be connected to molecular retrieval and editing [9], while multi-modal molecule-text modeling suggests a route to richer prompt interpretation [10]. A structured prompt could specify scaffold retention, reduced hERG liability, improved solubility, or acceptable synthetic accessibility, whereas a natural-language prompt could express the same goals in medicinal chemistry language. The key requirement is that prompt semantics map consistently to generation constraints and reward components.

Encoding the Prompt into the Model

The prompt can be tokenized and embedded as a conditioning signal for the molecular decoder. In one design, the embedding is prepended as learned context before the SMILES sequence, following the general logic of property-conditioned transformer generation [12]. In another design, cross-attention allows the decoder to attend to prompt tokens during molecular generation, an approach conceptually aligned with molecule-text representation learning [10]. SAFE-style molecular representations may further support controllable generation by organizing molecular sequences in a chemically meaningful form [11].

Disentangling Multiple Objectives

A central challenge is preventing one objective, such as potency or lipophilicity, from dominating all others during generation. Multi-objective conditional graph generation [5] and multi-parametric de novo design workflows [18] show that molecular

optimization should represent competing goals as separable but interacting constraints. Reinforcement-learning systems can incorporate multiple reward terms, but they must be designed so that synthetic feasibility or ADMET safety is not treated as an afterthought [2]. A prompt-conditioned model should therefore decompose the prompt into activity, ADMET, structure, and synthesis subgoals before recombining them into a balanced generation signal.

Foundation Model Architecture and Training

Pre-Training Phase

The pre-training phase would use a causal transformer trained to model the syntax and distribution of drug-like molecular strings. Randomized SMILES studies showed that alternative molecular string enumerations can improve the behavior of molecular generative models [21], while MolGPT demonstrated the feasibility of transformer-decoder molecular generation [8]. Larger chemical foundation model frameworks such as ChemBERTa-3 emphasize reusable pre-training infrastructure for molecular AI. Graph-oriented foundation models such as MolE also indicate that future architectures could combine sequence and graph representations for lead optimization [16].

Prompt-Conditional Fine-Tuning

Prompt-conditional fine-tuning would use molecules paired with property descriptions, structural annotations, or textual summaries of their desired profiles. Translation between molecules and natural language provides a foundation for pairing molecular structures with textual descriptions [9], while multi-modal molecule structure-text modeling extends this idea toward text-based molecular retrieval and editing [10]. Conditional molecular generation networks show how desired properties can be used to guide target-specific molecule design [19]. In the proposed system, the model would learn to generate molecules when given a prompt describing the bioactivity, ADMET, and synthesis profile to be satisfied.

Reinforcement Learning with Multi-Objective Reward

After conditional fine-tuning, reinforcement learning would align the generator with prompt satisfaction by rewarding molecules that jointly satisfy activity, ADMET, and synthesis constraints. Early deep reinforcement-learning systems established the use of molecular rewards for de novo design [1, 15], and REINVENT 2.0 provided a practical framework for reinforcement-learning-based molecular generation [22]. Multi-constraint transformer generation combined conditional modeling, knowledge distillation, and reinforcement learning for molecular design [12], while transformer policy-gradient methods further support reward-guided generation [17]. The reward in this EAI system would remain conceptual and should be evaluated carefully rather than presented as evidence of achieved lead optimization.

Multi-Objective Reward Design and Constrained Generation

Aggregating Multiple Property Scores

A prompt-conditioned lead optimization model would require a harmonized reward function that translates heterogeneous medicinal chemistry objectives into a shared decision signal. Bioactivity predictions, ADMET classifications, physicochemical filters, and synthetic feasibility estimates could be normalized conceptually so that no single score overwhelms the others, following the multi-constraint design logic of conditional transformer generation [12]. Reinforcement-learning systems for molecular generation have shown that reward terms can steer chemical output [1, 15], but in lead optimization the reward should remain adjustable by prompt context rather than fixed across all projects. Synthetic complexity measures such as SCScore [13] and retrosynthetic accessibility estimates [14] would be incorporated as synthesis-aware terms alongside biological and ADMET objectives.

Handling Tensions between Objectives

Lead optimization prompts often encode tensions, such as increasing polarity while preserving permeability, reducing lipophilicity while maintaining potency, or improving synthetic accessibility without removing essential recognition motifs. Multi-objective molecular design methods are useful here because they can treat a prompt as a request for a region of acceptable trade-offs rather than a single ideal molecule [5, 18]. Diversity-aware evaluation methods such as the Fréchet ChemNet Distance also highlight the importance of preserving chemically meaningful distributions rather than collapsing generation toward narrow score-maximizing structures [23]. A prompt-conditioned model should therefore use stochastic decoding, diversity preservation, and Pareto-aware selection to expose multiple plausible optimization paths.

Constraining the Output to a Chemical Series

In many lead optimization campaigns, the goal is not to invent an unrelated molecule but to preserve a privileged scaffold while modifying substituents, vectors, or physicochemical balance. Focused library generation from recurrent molecular models demonstrated how generative chemistry can remain close to a target chemical region [3], and later reaction-aware or constrained optimization frameworks extend this idea toward chemically guided analog exploration. A scaffold-matching constraint could be integrated into the reward so that generated molecules maintain the core series while exploring side-chain variation. Such constraints would help ensure that prompt-conditioned generation remains relevant to medicinal chemistry hypotheses rather than drifting into unrelated chemical space.

Model Evaluation And Validation In Lead Optimization

In-Silico Metrics

In-silico evaluation should assess whether generated molecules are valid, unique, novel, diverse, and consistent with the property profile requested in the prompt. GuacaMol formalized benchmark tasks for de novo molecular design [6], while MOSES provided a platform for comparing generative molecular models across distributional and structural criteria [7]. Fréchet ChemNet Distance adds a complementary view by measuring whether generated molecules resemble learned chemical distributions rather than merely satisfying isolated validity checks [23]. For prompt-conditioned lead optimization, these metrics should be interpreted alongside prompt satisfaction, scaffold retention, and synthesis-aware filtering.

Prospective Experimental Validation

Prospective validation would be essential because an apparently favorable molecule in silico may fail once synthesized and tested. The rapid identification of DDR1 kinase inhibitors using deep learning illustrates how generative design proposals can move toward experimental assessment in a real drug discovery context [24], while gene-expression-conditioned molecule generation shows that generative models can be connected to biologically grounded design objectives [25]. A prompt-conditioned lead optimization system should therefore be evaluated through retrospective case studies and carefully designed prospective campaigns, without assuming that model-generated candidates are automatically superior. Experimental validation would need to consider target activity, ADMET behavior, synthetic effort, and medicinal chemistry rationale together.

Integration Into The Medicinal Chemistry Workflow

Interactive Lead Optimization Sessions

In a practical medicinal chemistry workflow, the model would function as an interactive design assistant during project discussions. A chemist could revise the prompt to emphasize potency preservation, reduced clearance, improved solubility, or easier synthesis, and the system would generate a new set of candidate analogs for review. Platforms such as Chemistry42 illustrate how AI-driven molecular design can be organized into an optimization environment for medicinal chemistry users [26]. The prompt interface would be valuable because it translates expert intent into model conditioning without requiring the chemist to directly manipulate reward equations.

Complementing Structure-Based Design

Although the proposed system is primarily ligand- and prompt-driven, it could complement structure-based design by encoding binding-site constraints, pharmacophore requirements, or pose-derived hypotheses into the prompt. Foundation molecular generation approaches such as GP-MoLFormer suggest that generative models can be extended toward richer molecular design contexts. Molecule-text and molecular editing systems also support the idea that structural or mechanistic instructions can be represented as conditioning information [10]. In this setting, the model would not replace docking, free-energy analysis, or structural review, but would provide synthesis-aware analog ideas that can be filtered by those methods.

Evaluation Strategy

Quality of Generated Candidates

The quality of generated candidates should be evaluated by how consistently they satisfy prompt-specified criteria while remaining chemically plausible and medicinally interpretable. Benchmarking frameworks such as GuacaMol [6] and MOSES [7] provide useful starting points for assessing generated molecular sets, but prompt-conditioned lead optimization additionally requires measuring alignment between the user's requested profile and the generated analogs. Synthesis-aware criteria from SCScore [13] and RAscore [14] should be included so that feasibility is evaluated alongside bioactivity and ADMET predictions. The analysis should avoid overinterpreting aggregate scores and instead ask whether the candidate set offers actionable design hypotheses.

Table 2 provides a reward and validation framework for evaluating whether prompt-conditioned molecular generation produces chemically meaningful, synthesis-aware, and experimentally testable lead optimization hypotheses.

Table 2. Multi-Objective Reward and Validation Framework for Prompt-Conditioned Lead Optimization

Optimization dimension	Prompt-level expression	Candidate-level computational signal	Desired behavior in generated molecules	Trade-off tension	Validation requirement	Interpretation for medicinal chemist review
Target bioactivity	"Improve potency," "retain activity," "increase target engagement"	Predicted activity, QSAR score, docking-informed activity estimate, project-specific potency model	Molecules remain biologically plausible for the intended target	Higher potency may increase lipophilicity, molecular weight, or off-target risk	Retrospective activity benchmarks and prospective assay testing	Treat high predicted potency as a prioritization signal, not proof of activity
Scaffold or series retention	"Retain core scaffold," "modify only solvent-exposed	Scaffold match, substructure constraint, similarity window,	Generated analogs remain relevant to the existing lead series	Strong scaffold constraints may reduce novelty	Chemical-series review and medicinal chemistry rationale assessment	Useful when the goal is analog exploration rather

	vector," "preserve recognition motif"	pharmacophore preservation		and limit property improvement		than unrelated de novo discovery
Physicochemical balance	"Improve solubility," "reduce lipophilicity," "maintain permeability"	cLogP, polar surface area, hydrogen-bond counts, molecular weight, solubility predictors	Molecules move toward drug-like property ranges	Increased polarity may reduce permeability or potency	Distributional comparison against known leads and property-specific experimental testing	Should be evaluated as a balanced profile rather than isolated property improvement
ADMET liability reduction	"Reduce clearance risk," "avoid hERG liability," "improve metabolic stability"	ADMET classifiers, toxicity-alert models, metabolic stability predictors, off-target risk models	Candidate set avoids obvious developability liabilities	Safer ADMET profiles may require structural changes that weaken potency	External ADMET benchmarks and experimental ADMET panels	Strong ADMET predictions should guide triage but require empirical confirmation
Synthetic accessibility	"Keep synthesis simple," "avoid difficult transformations," "maintain accessible route"	SCScore-like complexity, RAscore-like retrosynthetic accessibility, reaction-feasibility estimate	Candidates appear chemically makeable and route-plausible	Simple synthesis may conflict with optimal potency or selectivity	Retrosynthetic planning, route review, reagent availability check, synthetic chemist assessment	Synthetic feasibility should be considered before committing to candidate synthesis
Diversity within the analog set	"Provide multiple optimization options," "avoid near-duplicate molecules"	Molecular fingerprints, clustering, pairwise similarity, distributional diversity metrics	Candidate set offers alternative chemical strategies	Too much diversity may move candidates away from the active series	Diversity analysis plus expert review of chemical plausibility	Diversity is valuable only when molecules remain relevant to the lead hypothesis
Prompt satisfaction	"Generate molecules matching the requested profile"	Alignment between requested constraints and predicted candidate properties	Outputs correspond to the user's stated design priorities	A molecule may satisfy one prompt component while violating another	Prompt-to-output audit across activity, ADMET, scaffold, and synthesis criteria	Enables chemists to judge whether the model understood the design intent
Pareto trade-off quality	"Balance potency, ADMET, and synthesis rather than maximize one score"	Pareto ranking, weighted utility, non-dominated candidate selection	Candidate set spans acceptable compromise solutions	Single-score optimization can hide clinically or chemically unacceptable compromises	Multi-objective benchmarking and sensitivity analysis of reward weights	Helps chemists compare candidates as trade-off options rather than absolute winners
Applicability-domain reliability	"Avoid unsupported predictions," "flag uncertain molecules"	Prediction confidence, similarity to training domain, uncertainty estimate	Molecules outside model confidence are flagged rather than over-ranked	Novel chemistry often has higher uncertainty	Applicability-domain analysis and uncertainty calibration	Uncertain molecules may still be interesting but should not be overinterpreted
Prospective usefulness	"Return candidates worth discussing, synthesizing, or testing"	Expert rating, synthesis decision, assay follow-through, design-cycle impact	Candidate set supports actionable medicinal chemistry decisions	In-silico scores may not translate into practical project value	Retrospective case studies followed by prospective lead-optimization campaigns	Final value depends on whether generated molecules improve real design decisions

Efficiency and Speed

Efficiency should be evaluated conceptually by comparing how a prompt-conditioned system changes the design process relative to conventional manual ideation and sequential computational filtering. REINVENT 2.0 illustrates how reinforcement-learning-based molecular generation can support iterative de novo design workflows [22], while Chemistry42 shows how AI-driven molecular optimization can be framed as a platform-level process for candidate proposal [26]. In this EAI setting, the relevant question is whether the model could reduce unproductive exploration and make design meetings more focused, not whether it achieves a predetermined numerical acceleration. Efficiency should therefore be judged through workflow fit, interpretability, and the usefulness of generated candidates for the next design cycle.

Medicinal Chemist Acceptance

Medicinal chemist acceptance should be assessed through expert review of generated molecules, prompt controllability, and perceived actionability of model suggestions. Conditional molecular generation systems such as CMGN show how target-specific property goals can be incorporated into generation [19], while Llamol illustrates dynamic multi-conditional generation for de novo molecular design [20]. However, acceptance will depend on whether chemists can understand why a molecule was proposed, adjust the prompt intuitively, and reject implausible candidates without disrupting the design workflow. A useful system should therefore support transparent prompt iteration and expert judgment rather than presenting model output as an unquestioned recommendation.

Limitations

Dependence on Predictive Model Accuracy

The reward function depends on predictive models for bioactivity, ADMET, and synthetic feasibility, and these models may be unreliable outside their applicable domains. Reinforcement learning can amplify weaknesses in the scoring functions when the generator discovers molecules that satisfy model artifacts rather than true medicinal chemistry objectives [1, 17]. Synthesis estimators such as SCScore [13] and RAscore [14] are valuable, but they remain proxies for real route design, reagent availability, yield, purification burden, and project-specific synthetic constraints. As a result, any generated molecule should be treated as a hypothesis requiring expert inspection and experimental confirmation.

Limited to Chemical Space Learned During Pre-Training

A molecular foundation model may struggle with scaffolds, elements, transformations, or property regimes that are poorly represented during pre-training. SMILES-based transformers such as MolGPT [8] and broader chemical foundation model frameworks such as ChemBERTa-3 depend on learned chemical distributions, which can support generalization but may also bias generation toward familiar motifs. Graph foundation models such as MolE [16] may help by representing molecular structure differently, yet they also require careful adaptation when used for novel chemistry. Additional fine-tuning, curated project data, and medicinal chemistry review would therefore remain necessary for unfamiliar lead series.

Conclusion

Prompt-conditioned molecular foundation models offer a conceptual route for integrating bioactivity, ADMET, and synthetic feasibility into a single lead optimization interface. Instead of treating each property as a separate modeling problem, the proposed system frames medicinal chemistry intent as a prompt that conditions molecular generation.

The main strength of this approach is intuitive control. A medicinal chemist could express desired changes in ordinary design language, while the model could generate chemically diverse analogs that reflect the requested balance of potency, developability, and synthesis.

Important challenges remain. Reward models must be reliable, scaffold constraints must be enforced without overrestricting creativity, and generated molecules must be validated through rigorous experimental workflows before they can influence project decisions.

Open-source tools and collaborative benchmarks will be important for evaluating prompt-driven generative chemistry in realistic lead optimization campaigns. Such benchmarks should emphasize chemical plausibility, synthesis awareness, expert interpretability, and prospective usefulness rather than isolated model scores.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Olivecrona M, Blaschke T, Engkvist O, Chen H. Molecular de-novo design through deep reinforcement learning. J Cheminform. 2017;9(1):48.
2. Ståhl N, Falkman G, Karlsson A, Mathiason G, Bostrom J. Deep reinforcement learning for multiparameter optimization in de novo drug design. J Chem Inf Model. 2019;59(7):3166-76.
3. Segler MH, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120-31.
4. Gómez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76.
5. Li Y, Zhang L, Liu Z. Multi-objective de novo drug design with conditional graph generative model. J Cheminform. 2018;10(1):33.
6. Brown N, Fiscato M, Segler MH, Vaucher AC. GuacaMol: benchmarking models for de novo molecular design. J Chem Inf Model. 2019;59(3):1096-108.
7. Polykovskiy D, Zhebrak A, Sanchez-Lengeling B, Golovanov S, Tatanov O, Belyaev S, et al. Molecular sets (MOSES): a benchmarking platform for molecular generation models. Front Pharmacol. 2020;11:565644.
8. Bagal V, Aggarwal R, Vinod PK, Priyakumar UD. MolGPT: molecular generation using a transformer-decoder model. J Chem Inf Model. 2021;62(9):2064-76.
9. Edwards C, Lai T, Ros K, Honke G, Cho K, Ji H. Translation between molecules and natural language. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2022; p. 375-413.

10. Liu S, Nie W, Wang C, Lu J, Qiao Z, Liu L, et al. Multi-modal molecule structure–text model for text-based retrieval and editing. Nat Mach Intell. 2023;5(12):1447-57.
11. Noutahi E, Gabellini C, Craig M, Lim JS, Tossou P. Gotta be SAFE: a new framework for molecular design. Digital Discovery. 2024;3(4):796-804.
12. Wang J, Hsieh CY, Wang M, Wang X, Wu Z, Jiang D, et al. Multi-constraint molecular generation based on conditional transformer, knowledge distillation and reinforcement learning. Nat Mach Intell. 2021;3(10):914-22.
13. Coley CW, Rogers L, Green WH, Jensen KF. SCScore: synthetic complexity learned from a reaction corpus. J Chem Inf Model. 2018;58(2):252-61.
14. Thakkar A, Chadimová V, Bjerrum EJ, Engkvist O, Reymond JL. Retrosynthetic accessibility score (RAscore)–rapid machine learned synthesizability classification from AI driven retrosynthetic planning. Chem Sci. 2021;12(9):3339-49.
15. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):eaap7885.
16. Méndez-Lucio O, Nicolaou CA, Earnshaw B. MolE: a foundation model for molecular graphs using disentangled attention. Nat Commun. 2024;15(1):9431.
17. Mazuz E, Shtar G, Shapira B, Rokach L. Molecule generation using transformers and policy gradient reinforcement learning. Sci Rep. 2023;13(1):8799.
18. Perron Q, Mirguet O, Tajmouati H, Skiredj A, Rojas A, Gohier A, et al. Deep generative models for ligand-based de novo design applied to multi-parametric optimization. J Comput Chem. 2022;43(10):692-703.
19. Yang M, Sun H, Liu X, Xue X, Deng Y, Wang X. CMGN: a conditional molecular generation net to design target-specific molecules with desired properties. Brief Bioinform. 2023;24(4):bbad185.
20. Dobberstein N, Maass A, Hamaekers J. Llamol: a dynamic multi-conditional generative transformer for de novo molecular design. J Cheminform. 2024;16(1):73.
21. Arús-Pous J, Johansson SV, Prykhodko O, Bjerrum EJ, Tyrchan C, Reymond JL, et al. Randomized SMILES strings improve the quality of molecular generative models. J Cheminform. 2019;11(1):71.
22. Blaschke T, Arús-Pous J, Chen H, Margreitter C, Tyrchan C, Engkvist O, et al. REINVENT 2.0: an AI tool for de novo drug design. J Chem Inf Model. 2020;60(12):5918-22.
23. Preuer K, Renz P, Unterthiner T, Hochreiter S, Klambauer G. Fréchet ChemNet distance: a metric for generative models for molecules in drug discovery. J Chem Inf Model. 2018;58(9):1736-41.
24. Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038-40.
25. Méndez-Lucio O, Baillif B, Clevert DA, Rouquié D, Wichard J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. Nat Commun. 2020;11(1):10.
26. Ivanenkov YA, Polykovskiy D, Bezrukov D, Zagribelnyy B, Aladinskiy V, Kamya P, et al. Chemistry42: an AI-driven platform for molecular design and optimization. J Chem Inf Model. 2023;63(3):695-701.