AUTONOMOUS DISCOVERY OF TRAINING DISTRIBUTIONS FOR CONTROLLABLE MOLECULAR GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Molecular generative models are typically pretrained on fixed drug-like datasets, implicitly constraining the structural support of the learned representation and limiting downstream adaptability. We investigate whether pretraining distribution support acts as a controllable lever on generative expressivity, and whether expanding this support through structurally distant molecules improves downstream diversity–activity trade-offs. We instantiate this hypothesis using a large language model (LLM) as an autonomous research agent and a lightweight autoregressive Transformer as the molecular generator. The agent iteratively proposes molecular pretraining mixtures, the generator is retrained under fixed model capacity and compute budget, and performance is evaluated on the task of spleen tyrosine kinase (SYK) inhibitor generation and optimization using a joint objective combining scaffold entropy and predicted pIC_{50} after fine-tuning on known SYK inhibitors. Across multiple iterations, the search identifies boundary configurations composed of molecules maximally distant from canonical drug-like datasets, yielding higher diversity–activity scores than conventional baselines. The best-performing configuration improves the joint diversity–activity objective by 1.6% relative to the baseline, driven primarily by increased scaffold entropy while maintaining competitive predicted affinity. These results suggest that training distribution discovery can serve as a data-centric mechanism for improving controllable molecular generation under constrained model capacity.

1 INTRODUCTION

Deep generative models have become a central paradigm for *de novo* molecular design. Sequence-based and graph-based architectures, including recurrent neural networks, transformer decoders, and diffusion models, have demonstrated strong performance in learning chemical distributions and generating candidate molecules (Brown et al., 2019; Polykovskiy et al., 2020; Bagal et al., 2021; Hoogeboom et al., 2022). More recently, large-scale molecular language models have emphasized architectural scaling and massive pretraining as mechanisms for expanding accessible chemical space (Ross et al., 2022; Brinkmann et al., 2022; Zhang et al., 2025).

Despite these advances, one fundamental component of generative modeling remains largely underexplored: the structure of the pretraining distribution itself. In most molecular generative pipelines, the training dataset is treated as fixed, typically a curated drug-like subset such as ChEMBL (Davies et al., 2015) or ZINC (Irwin et al., 2020), and model performance is optimized by adjusting architecture, scaling model capacity, or modifying conditioning objectives.

This architectural emphasis contrasts with the broader movement toward data-centric artificial intelligence (AI), which argues that dataset composition and distribution design can be as influential as model scaling (Ng, 2021; Wang et al., 2020). While dataset curation and curriculum strategies have been studied in supervised learning (Bengio et al., 2009), systematic optimization of pretraining distributions for molecular generative models remains largely underexplored.

Inspired by recent proposals for LLM-driven autonomous experimentation (Shinn et al., 2023; Wang et al., 2023; Karpathy, 2026), we adopt a data-centric perspective on controllable molecular generation and ask whether the pretraining distribution itself can be treated as an object of optimization.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We hypothesize that the structural support of the pretraining distribution acts as a controllable lever on generative expressivity.

Our contributions are threefold:

- We formalize pretraining distribution discovery for molecular generative models as a closed-loop optimization problem over dataset mixtures. Specifically, we investigate whether expanding structural support through the inclusion of molecules that are maximally distant from canonical drug-like datasets improves downstream adaptability under constrained model capacity.
- We introduce an autonomous closed-loop framework for discovering training distributions. Our system employs a large language model (LLM) as a history-aware research agent that iteratively proposes, evaluates, and refines dataset mixture hypotheses using scalar reward feedback. Crucially, all model architecture and optimization parameters are fixed, isolating the effect of training distribution composition.
- Across experiments on optimization of inhibitors for spleen tyrosine kinase (SYK), a clinically important therapeutic target with an ongoing need for novel inhibitors, the autonomous search identifies a boundary solution corresponding to maximal structural distance from ChEMBL. The best-performing configuration improves a joint diversity–activity objective by 1.6% relative to baseline, driven primarily by increased scaffold entropy rather than increased predicted affinity. We make all the agent instructions and code for this study publicly available at `https://anonymous.4open.science/r/molecular_autoresearch-F0F0`.

2 Related Work

Molecular Generative Modeling. Deep generative models for *de novo* design include SMILES/graph generators based on Transformers and diffusion models, with standard evaluation suites such as MOSES and GuacaMol (Polykovskiy et al., 2020; Brown et al., 2019; Bagal et al., 2021; Hoogeboom et al., 2022). Most work improves generation via architecture, scaling, or objectives while treating the pretraining corpus as fixed. In contrast, we study *pretraining distribution composition* itself as the optimization variable under a frozen model/training setup.

Target-Specific Molecular Generation. Target adaptation is commonly achieved via fine-tuning, conditional generation, or RL-style optimization toward property/docking/QSAR objectives (Winter et al., 2019; Popova et al., 2018; Zhou et al., 2019; Corso et al., 2023). These methods optimize behavior for a given dataset/target, but typically do not isolate how changing the *pretraining* data distribution affects downstream diversity–activity trade-offs under fixed capacity.

Scaling and data-centric perspectives. Large-scale chemical language models emphasize architectural scaling and massive corpora to expand accessible chemical space (Ross et al., 2022; Brinkmann et al., 2022; Zhang et al., 2025). Separately, data-centric AI highlights that dataset design can be as important as model changes (Ng, 2021; Wang et al., 2020; Bengio et al., 2009). We connect these viewpoints by optimizing mixtures/curation policies of molecular pretraining data while keeping the model and training budget fixed.

Autonomous agents for closed-loop optimization. LLM-based agents can iteratively propose actions, observe feedback, and refine hypotheses in closed loops (Shinn et al., 2023; Wang et al., 2023; Karpathy, 2026). We adapt this paradigm to molecular generator pretraining by using an LLM agent to search over pretraining mixtures and curation operators, evaluated via a downstream diversity–activity objective.

3 Methodology

We propose a two-layer framework for the autonomous discovery of training distributions in controllable molecular generation. As illustrated in Figure 1, the framework consists of an Instruction Generation Layer, responsible for transforming a human research hypothesis into a structured,

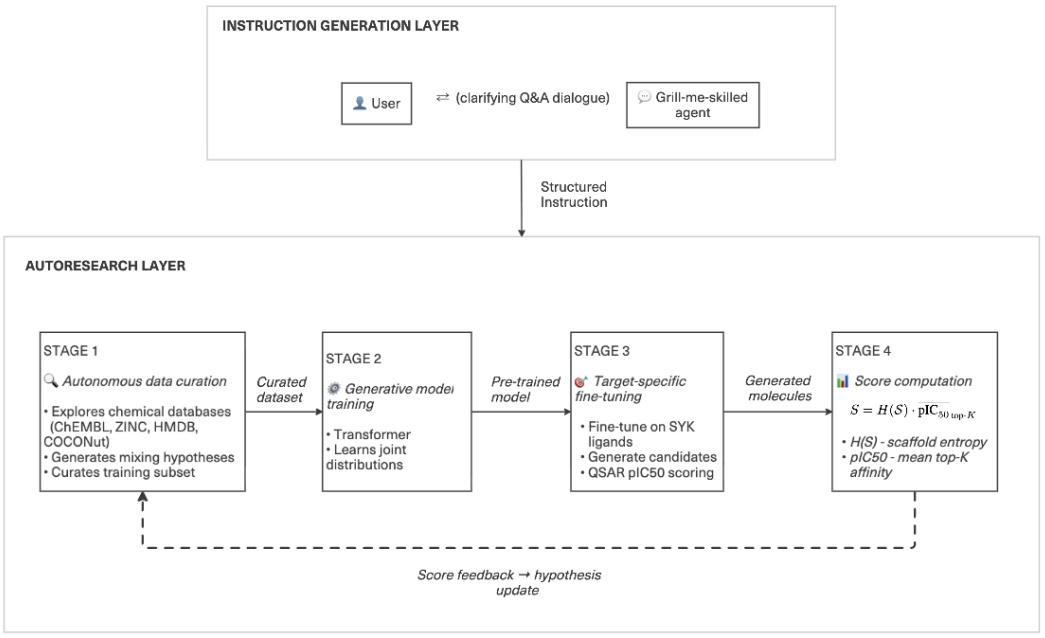


Figure 1: **Overview of the proposed autonomous training distribution discovery pipeline.** The INSTRUCTION GENERATION LAYER (top) transforms a human-formulated research hypothesis into a structured instruction via an interactive clarification dialogue with an LLM-based Agent using the *Grill-me* skill (Pocock, 2023). The AUTORESEARCH LAYER (bottom) iteratively executes four stages: (1) autonomous data curation through dataset mixing hypotheses, (2) pretraining of a Transformer-based generative model, (3) target-specific fine-tuning and candidate generation, and (4) composite score computation combining scaffold entropy and predicted activity. The dashed arrow denotes feedback from the score to the hypothesis generator, forming a closed optimization loop.

machine-readable specification, and an AUTORESEARCH LAYER, which autonomously explores hypotheses through a closed feedback loop.

3.1 PROBLEM FORMULATION

Let $\mathcal{M}$ denote the space of valid molecules represented as SMILES strings. We assume access to a finite set of molecular data sources

$$\mathcal{D} = \{D_1, \ldots, D_n\},$$

including ChEMBL, ZINC, and other chemical databases with varying structural properties.

Given a target protein $\mathcal{T}$ (SYK kinase in our experiments) and a downstream evaluation functional $\mathcal{J}$, our objective is to discover a pretraining distribution p_{train} such that a generative model pretrained on samples from p_{train} and subsequently fine-tuned on $\mathcal{T}$ maximizes $\mathcal{J}$:

$$p^*_{\text{train}} = \arg\max_{p_{\text{train}} \in \mathcal{P}_{\text{mix}}} \mathcal{J}\left(f_{\theta^*(p_{\text{train}})}, \mathcal{T}\right), \tag{1}$$

where $f_{\theta^*(p_{\text{train}})}$ denotes a generative model whose parameters are obtained by: pretraining on samples from p_{train} and fine-tuning on known ligands of $\mathcal{T}$.

Rather than optimizing over arbitrary distributions in $\Delta(\mathcal{M})$, the search space is restricted to finite mixtures of available datasets:

$$\mathcal{P}_{\text{mix}} = \left\{ \sum_{j \in \mathcal{I}} \alpha_j p_j \;\middle|\; \mathcal{I} \subseteq \{1, \ldots, n\}, \quad \sum_{j \in \mathcal{I}} \alpha_j = 1,\ \alpha_j \geq 0 \right\}. \tag{2}$$

Here, $D_j \in \mathcal{D}$ denotes the j-th molecular dataset, and $p_j \in \Delta(\mathcal{M})$ is the empirical distribution induced by D_j. The index set $\mathcal{I}$ specifies the subset of datasets included in the mixture, and α_j are non-negative mixture coefficients that sum to one. Thus, each $p_{\text{train}} \in \mathcal{P}_{\text{mix}}$ is a convex combination of empirical dataset distributions. In practice, mixtures are instantiated as fixed-size curated subsets with varying proportions and optional filtering predicates.

3.2 Datasets and Preprocessing

Data sources. We consider a collection of publicly available molecular datasets with heterogeneous structural and physicochemical properties:

- **ChEMBL**: bioactive molecules with annotated activities. (Davies et al., 2015)
- **ZINC**: commercially available compounds. (Irwin et al., 2020)
- **COCONUT**: collection of open natural products (Sorokina & Steinbeck, 2021)
- **HMDB**: a curated database of endogenous and exogenous human metabolites. (Wishart et al., 2022)

Standardization. For each dataset, only SMILES representations of molecules are retained. All molecules are processed using RDKit framework. SMILES strings are parsed into molecular objects with built-in sanitization and valence validation. Invalid or unparsable structures are discarded. Canonical SMILES representations are generated for all valid molecules. Molecules exceeding a maximum SMILES length of 128 characters are removed. Duplicate molecules are eliminated based on canonical SMILES identity.

Tokenization. SMILES strings are tokenized using a fixed rule-based tokenizer with greedy longest-match parsing over a predefined vocabulary of chemically meaningful tokens, following common practice in SMILES-based generative modeling (Schwaller et al., 2019; Olivecrona et al., 2017). The vocabulary is fixed across all datasets and constructed from frequently occurring bracketed groups in ChEMBL, ZINC, COCONUT, and HMDB. Rare or unseen bracketed tokens fall back to a single `<unk>` symbol. Four special tokens are used: `<pad>`, `<bos>`, `<eos>`, and `<unk>`. The total vocabulary size is $|\mathcal{V}| = 165$ tokens (including special symbols). All sequences are truncated or padded to a maximum length of 128 tokens. During autonomous mixture search, curated datasets D_t are constructed as fixed-size samples (500,000 molecules) drawn according to the mixture coefficients $\{\alpha_j\}$ specified by hypothesis h_t.

Chemical space overlap metrics. To further characterize structural differences between data sources, we computed two complementary chemical space overlap metrics between ChEMBL and ZINC. From the standardized datasets, we randomly sampled 28,548 molecules from ChEMBL (1 % sample from full dataset) and 7,461,461 molecules from ZINC without replacement using a fixed random seed. These subsets were used exclusively for cross-dataset similarity analysis.

For each molecule, we generated Morgan fingerprints (ECFP4) with radius $r = 2$ and 2048 bits. In addition, Bemis–Murcko scaffolds were extracted and converted to canonical non-isomeric SMILES.

For every molecule z in the ZINC sample, we computed its cumulative Tanimoto similarity to all sampled ChEMBL molecules:

$$\mathrm{TSum}(z) = \sum_{c \in \mathcal{C}} \mathrm{Tanimoto}(\mathrm{FP}(z), \mathrm{FP}(c)), \tag{3}$$

where $\mathcal{C}$ denotes the 28,548-molecule ChEMBL subset. This scalar quantity measures the global fingerprint-level proximity of a molecule to the bioactive chemical space represented by ChEMBL.

To capture structural core overlap, we additionally computed the number of exact Bemis–Murcko scaffold matches:

$$\mathrm{SCount}(z) = \sum_{c \in \mathcal{C}} \mathbf{1}\left[\mathrm{Scaffold}(z) = \mathrm{Scaffold}(c)\right]. \tag{4}$$

The two metrics provide complementary signals: TSum reflects continuous chemical similarity, while SCount quantifies discrete scaffold-level overlap. Together, they characterize the degree of structural alignment or novelty between the two chemical spaces.

3.3 Instruction Generation Layer

The process begins with a human-formulated research hypothesis. Rather than directly specifying dataset mixtures or evaluation criteria, the user provides a high-level intent (e.g., exploring whether non-druglike molecules improve adaptability for SYK inhibitors).

To transform this informal hypothesis into a structured executable specification, we employ a *Grill-me-skilled agent* — an LLM-based clarification module that engages the user in an iterative question–answer dialog (Pocock, 2023). Given an initial prompt u_0, the Grill-me Agent generates targeted clarifying questions addressing target protein definition, acceptable chemical novelty, evaluation metrics, search constraints and computational budget, and dataset inclusion preferences.

Collecting the responses, the agent produces a structured instruction, which encodes the research goal in a machine-readable format. This layer explicitly separates *hypothesis formulation* (human + Grill-me) from *hypothesis execution* (autonomous search), improving reproducibility and reducing ambiguity.

3.4 Autoresearch Layer

The closed-loop refinement process is inspired by recent proposals for autonomous experiment-driven model improvement (Karpathy, 2026), but is specialized here for structured molecular data mixture optimization. The autonomous hypothesis refinement process is implemented using a Qwen-3.5-27B large language model, which operates as a history-aware research agent under explicit constraints (see details in Appendix section B). At each iteration t, the agent proposes a single hypothesis h_t describing a candidate mixture.

Stage 1: Autonomous data curation. Each hypothesis is defined as:

$$h_t = (\mathcal{S}_t, \{\alpha_j\}_{j \in \mathcal{S}_t}, \phi_t)\,, \tag{5}$$

where $\mathcal{S}_t$ is a subset of datasets, α_j are mixing coefficients, ϕ_t is an optional filtering predicate.

The agent constructs a curated dataset D_t of fixed size according to h_t.

In addition to selecting mixture coefficients, the agent proposes structured sampling strategies that determine how samples are drawn from the induced mixture. These include:

- Random sampling,
- Scaffold-first selection,
- Property-balanced sampling,
- Distance-based diversification from a reference set,
- Non-drug-like enrichment,
- High predicted activity prioritization,
- Class-balanced sampling.

These operators induce controlled shifts in structural and physicochemical support, enabling exploration of diversity expansion and target-biased enrichment within a fixed training budget. Each strategy corresponds to a distinct inductive bias over chemical space, allowing the agent to probe qualitatively different regions of the mixture landscape. Importantly, all interventions are restricted to data selection, leaving model architecture, tokenization, and optimization unchanged.

Although the search space is defined over convex mixtures of available datasets, the optimization is not constrained to interior solutions. In practice, the autonomous search may converge to boundary configurations of the mixture space, corresponding to pure or near-pure data sources. Such boundary

solutions reflect the agent's assessment of maximal structural support expansion under the fixed training budget.

Stage 2: Generative model training. We employ a lightweight GPT-style autoregressive Transformer with rotary positional embeddings (RoPE) and RMS normalization as the molecular generator. The model is trained to predict next-token probabilities over SMILES sequences using standard cross-entropy loss. All architectural parameters (number of layers, embedding dimension, attention heads) are fixed across experiments (see details in Appendix section B).

Stage 3: Target-specific fine-tuning. The pretrained model is fine-tuned on known ligands of $\mathcal{T}$, producing $f_{\theta_t^*}$. We generate N candidate molecules and score them with a QSAR predictor $\hat{y}(x, \mathcal{T})$ estimating pIC_{50}. The model and known ligands were taken from (Zavadskaya et al., 2025).

Stage 4: Composite score. To jointly optimize structural diversity and predicted potency, we define:

$$S(h_t) = H(\mathcal{S}_t) \cdot \overline{\text{pIC}_{50}}_{\text{top-}K}(h_t), \tag{6}$$

where H is scaffold entropy and $\overline{\text{pIC}_{50}}_{\text{top-}K}$ is the mean predicted affinity of the top-K candidates.

We adopt a multiplicative objective rather than an additive form to prevent degenerate solutions. If either diversity or activity collapses, the product is reduced, discouraging optimization of one factor at the expense of the other.

While additional quality metrics such as validity are monitored, the objective explicitly prioritizes high-affinity top-ranked candidates. Since downstream molecular discovery focuses on the most promising molecules rather than the full generated distribution, we optimize performance at the top of the ranking rather than global generation validity.

Hypothesis refinement. The agent operates sequentially. At iteration t, it receives the full history $\{h_1, \ldots, h_{t-1}\}$, all associated metrics (entropy, activity, validity), the composite score $S(h_{t-1})$, and the baseline performance. Although all metrics are visible, the instruction explicitly designates S as the primary optimization signal. The agent compares $S(h_{t-1})$ to the baseline and prior experiments, and proposes a new hypothesis h_t by modifying mixture proportions or dataset selection. This process continues for a fixed computational budget.

To isolate the effect of training distribution, all model-related parameters including model architecture and tokenizer, optimization hyperparameters, pretraining and fine-tuning schedules, random seed, and training time budget are strictly fixed across experiments. The only variable controlled by the Autonomous Research Agent is the composition of the pretraining dataset.

4 Results

4.1 Autonomous Search Discovers a Boundary Solution in Mixture Space

We evaluate whether an LLM-driven autonomous research loop can identify pretraining distributions that improve controllable molecular generation under a strictly fixed model configuration and training budget. Across all experiments, the agent was restricted to modifying only the composition of the pretraining dataset. Model architecture, tokenizer, optimization hyperparameters, random seed, and training schedule were held constant.

The left panel of Figure 2 visualizes the diversity–activity landscape across all explored pretraining configurations (see details in Appendix section C). The baseline (yellow) corresponds to pretraining on 500K randomly sampled ChEMBL molecules. Configurations outperforming the baseline (green) consistently exhibit increased scaffold entropy.

The highest-performing configuration (red) corresponds to pretraining exclusively on ZINC molecules selected to be maximally distant from ChEMBL in fingerprint space. Rather than converging to an interior mixture of drug-like and non-drug-like data, the autonomous search identifies

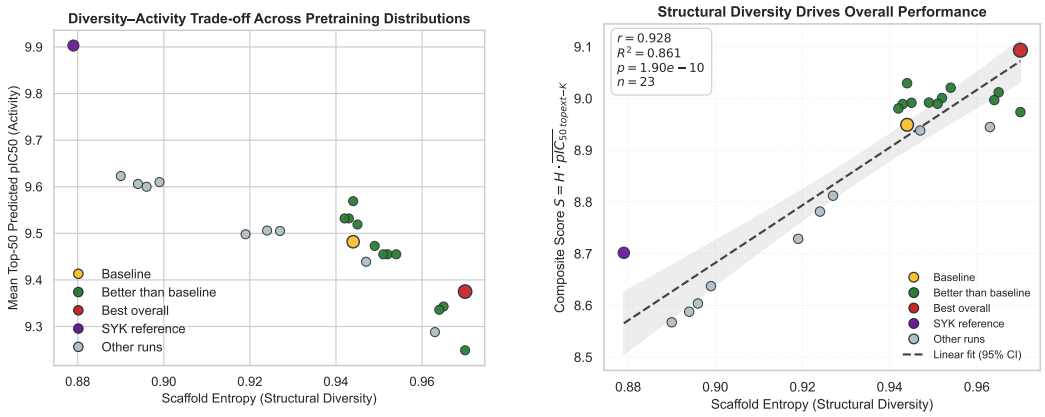


Figure 2: **Left:** Diversity–activity landscape across all explored pretraining distributions. Yellow: baseline (500K ChEMBL random). Green: configurations outperforming baseline. Red: best-performing configuration. Purple: SYK inhibitor reference set. The highest score is achieved by a boundary solution corresponding to maximal structural distance from canonical drug-like support. **Right:** Composite score as a function of scaffold entropy. The dashed line indicates a linear fit with 95% confidence interval. Structural diversity strongly correlates with overall performance ($R^2 = 0.861$).

a boundary solution of the mixture space, emphasizing maximal structural support expansion. This configuration achieves the highest composite score (+1.6% relative to baseline), driven primarily by increased scaffold entropy (0.970 vs 0.944), while maintaining competitive predicted activity.

Statistical evaluation confirms that this advantage is reproducible: the selected configuration achieves a higher composite score than the baseline (9.025 ± 0.033 vs. 8.956 ± 0.020, $p < 0.05$ across 5 seeds). The improvement is primarily associated with higher scaffold entropy (0.952 ± 0.001 vs. 0.941 ± 0.003, $p < 0.05$ across 5 seeds), whereas predicted pIC_{50} and validity does not differ significantly (9.478 ± 0.06 vs. 9.515 ± 0.028, $p = 0.29$ for pIC_{50} and 0.617 ± 0.06 vs. 0.665 ± 0.01, $p = 0.16$ for validity), supporting the conclusion that the gain arises from increased structural diversity rather than improved predicted activity (details can be seen in Appendix C).

4.2 Improvements Are Diversity-Driven Rather Than Activity-Driven

Although mean predicted pIC_{50} varies only moderately across experiments (9.29–9.62), scaffold entropy exhibits substantial variation and strongly influences the overall objective.

To further quantify this relationship, the right panel Figure 2 shows the composite score as a function of scaffold entropy. Structural diversity alone explains a large fraction of the variance in performance ($R^2 = 0.861$), indicating a strong linear dependence between entropy and score.

Importantly, the strong linear dependence between entropy and composite score suggests that, under constrained model capacity, expanding structural support may be more impactful than further architectural scaling.

4.3 Autonomous Refinement Behavior

The autonomous refinement process provides additional insight into how the boundary solution was discovered. Figure 3 shows the progression of the composite score across search iterations. The agent systematically evaluated moderate enrichment strategies, including small additions of natural products and balanced ChEMBL–ZINC mixtures, which failed to improve the objective.

Notably, the highest-performing configuration was identified relatively early in the search process. Subsequent iterations explored alternative mixture configurations but did not surpass this boundary solution.

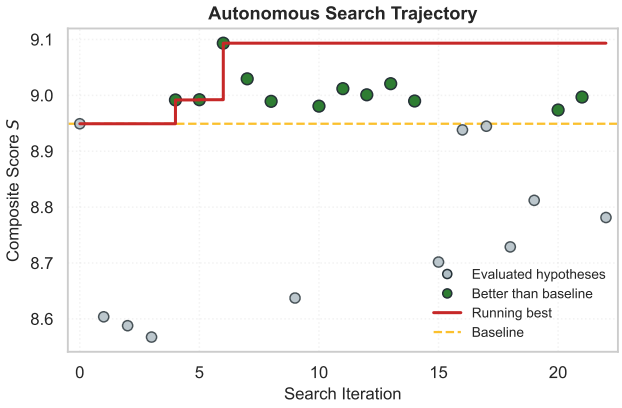


Figure 3: Autonomous search trajectory. Points denote evaluated hypotheses; the solid line indicates the running best score. The best-performing configuration is identified early, after which alternative mixtures are explored without exceeding this value.

This behavior suggests that maximal structural support expansion constitutes a dominant direction in mixture space under the fixed training budget, and that the autonomous agent is capable of rapidly identifying this direction using scalar reward feedback alone.

4.4 Top-K Optimization and Validity Trade-Off

While global validity decreases for highly diverse configurations, our objective explicitly prioritizes top-ranked molecules. Since drug discovery focuses on identifying a small number of high-affinity candidates, optimization at the top-K level provides a more relevant evaluation criterion than maximizing overall generation validity.

The results therefore indicate that expanding structural support during pretraining enhances downstream adaptability by broadening the representational coverage of chemical space, rather than by directly increasing predicted affinity across the entire generated distribution.

5 Discussion

5.1 Training Distribution as a Control Lever

Our results demonstrate that, under fixed model capacity and training budget, the choice of pretraining distribution exerts a substantial influence on downstream controllable molecular generation. Rather than converging to a balanced mixture of drug-like and non-drug-like datasets, the autonomous search identifies a boundary solution corresponding to maximal structural distance from canonical drug-like support.

This finding suggests that expanding structural support during pretraining increases the effective representational coverage of chemical space. Importantly, the strong linear dependence between scaffold entropy and composite score indicates that structural diversity is the dominant driver of performance in this setting.

Notably, the objective emphasizes top-ranked candidate molecules rather than global validity. While highly diverse pretraining distributions may reduce overall generation validity, drug discovery pipelines ultimately focus on identifying a small number of high-affinity candidates. Thus, optimizing performance at the top-K level provides a more task-aligned evaluation criterion.

These observations support a data-centric perspective: when model capacity is constrained, modifying the training distribution may yield larger gains than architectural scaling alone.

5.2 Autonomous Hypothesis Refinement

The autonomous agent operates using scalar reward feedback and full historical context. Despite lacking gradient access to the objective, the agent rapidly identifies structural support expansion as a dominant direction in mixture space.

Notably, the best-performing configuration is discovered early in the search process, after which subsequent iterations explore alternative hypotheses without surpassing this solution. This behavior suggests that the structural signal in the reward landscape is sufficiently strong to guide sequential hypothesis refinement without explicit optimization machinery.

However, the search remains heuristic and LLM-driven rather than formally optimized. Future work could integrate more principled Bayesian or evolutionary search strategies while preserving the autonomous closed-loop framework.

5.3 Limitations

Several limitations should be noted. First, the autonomous refinement process relies on large language model reasoning within a finite context window. As the number of experiments grows, maintaining full historical awareness may become challenging without explicit memory compression or structured summarization mechanisms.

Second, although the search space was explicitly constrained to dataset mixtures, the LLM-based agent occasionally proposed modifications outside the permitted configuration space. These proposals were programmatically rejected, but this behavior highlights the importance of robust constraint enforcement in autonomous scientific systems.

Third, in rare cases the agent selected non-optimal fine-tuning epochs for comparison when interpreting training logs. While these instances were filtered during evaluation, they illustrate the sensitivity of LLM-driven research agents to log formatting and reward signal presentation.

These limitations underscore that the framework is not a fully autonomous scientific discovery system, but rather a structured closed-loop research assistant operating under explicit guardrails.

6 Conclusion

We introduced a data-centric framework for autonomous discovery of training distributions in controllable molecular generation. Rather than modifying model architecture or scaling capacity, we treated the pretraining distribution itself as an optimization variable and implemented a closed-loop, history-aware refinement process over dataset mixtures.

Across experiments on SYK kinase optimization under fixed model capacity, the autonomous search converged to a boundary solution corresponding to maximal structural distance from canonical drug-like support. Performance gains in the composite diversity–activity objective were primarily driven by increased scaffold entropy, highlighting structural diversity as a key explanatory factor.

Overall, our results suggest that, under constrained capacity, expanding structural support during pretraining can yield improvements comparable to architectural scaling, positioning training distribution design as an underexplored axis of optimization in generative modeling.

Future work should investigate generalization across multiple protein targets and extend the framework beyond chemistry to other scientific domains. Agent-based optimization strategies may substantially accelerate research workflows, but this requires systematic study of how to control, constrain, and reliably tune autonomous agents. Understanding how training data support shapes learned representations across domains remains an important theoretical direction.

7 AI Use Statement

Generative AI tools were used to assist with the writing and editing of the manuscript, as well as to conduct the core experiments involving LLM agent. All AI-generated content, experimental

procedures, outputs, and conclusions were reviewed and verified by the authors. The authors take full responsibility for the content of the paper and the reported results.

8 Reproducibility Statement

We make the code, expert annotations, prompt templates, and evaluation scripts required to reproduce the main experiments publicly available through the anonymized repository `https://anonymous.4open.science/r/molecular_autoresearch-F0F0/`. The appendix documents the dataset description (Appendix A), additional experimental data (Appendix B) and extended results (Appendix C).

References

Viraj Bagal, Rohan Aggarwal, Piyush Vinod, and U. Deva Priyakumar. Molgpt: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling*, 61(12): 5905–5914, 2021.

Yoshua Bengio et al. Curriculum learning. In *ICML*, 2009.

Alexander Brinkmann et al. Chemgpt: Large-scale self-supervised pretraining for molecular property prediction. *NeurIPS Workshop*, 2022.

Nathan Brown, Matteo Fiscato, Marwin Segler, and Alain Vaucher. Guacamol: Benchmarking models for de novo molecular design. *Journal of Chemical Information and Modeling*, 59(3): 1096–1108, 2019.

Gabriele Corso, Hannes Stärk, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Diffdock: Diffusion steps, twists, and turns for molecular docking. In *ICLR*, 2023.

Mark Davies, Michał Nowotka, George Papadatos, and et al. Chembl web services: streamlining access to drug discovery data. *Nucleic Acids Research*, 43(W1):W612–W620, 2015.

Emiel Hoogeboom, Victor Satorras, et al. Equivariant diffusion for molecule generation. *ICML*, 2022.

John J. Irwin, Ke Tang, John Young, and et al. Zinc20—a free ultralarge-scale chemical database for ligand discovery. *Journal of Chemical Information and Modeling*, 60(12):6065–6073, 2020.

Andrej Karpathy. Autonomous research with llms. `https://github.com/karpathy/autoresearch`, 2026. Accessed: 2026-07-26.

Andrew Ng. A data-centric approach to ai. Landing AI, 2021.

Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1):48, 2017.

Matt Pocock. Grill-me skill. `https://github.com/mattpocock/skills/blob/main/skills/productivity/grill-me/SKILL.md`, 2023. Accessed: 2026-07-26.

Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, et al. Molecular sets (moses): A benchmarking platform for molecular generation models. *Frontiers in Pharmacology*, 11:565644, 2020.

Mariya Popova, Olexandr Isayev, and Alexander Tropsha. Deep reinforcement learning for de novo drug design. *Science Advances*, 4(7):eaap7885, 2018.

Jared Ross et al. Molecular representation learning with language models. *Nature Machine Intelligence*, 4, 2022.

Philippe Schwaller, Théophile Gaudin, David Lanyi, Costas Bekas, and Teodoro Laino. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS Central Science*, 5(9):1572–1583, 2019. doi: 10.1021/acscentsci.9b00576.

Noah Shinn et al. Reflexion: Language agents with verbal reinforcement learning. In *NeurIPS*, 2023.

Maria Sorokina and Christoph Steinbeck. Coconut online: Collection of open natural products database. *Journal of Cheminformatics*, 13(1):2, 2021.

Tongzhou Wang et al. Dataset distillation. In *NeurIPS*, 2020.

Zi Wang et al. Voyager: An open-ended embodied agent with large language models. *arXiv*, 2023.

Robin Winter, Floriane Montanari, Frank Noé, and Djork-Arné Clevert. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chemical Science*, 10(6):1692–1701, 2019.

David S. Wishart, An Chi Guo, Edgar Oler, and et al. Hmdb 5.0: the human metabolome database for 2022. *Nucleic Acids Research*, 50(D1):D622–D631, 2022. doi: 10.1093/nar/gkab1062.

Maria Zavadskaya, Anastasiia Orlova, Andrei Dmitrenko, et al. Integrating qsar modelling with reinforcement learning for syk inhibitor discovery. *Journal of Cheminformatics*, 17(1):52, 2025. doi: 10.1186/s13321-025-00998-2.

X. Zhang, Y. Li, H. Wang, et al. Expanding chemical space: Developing a compound generative pre-trained transformer for de novo drug design. *bioRxiv*, 2025. doi: 10.1101/2025.01.24.634665. URL `https://www.biorxiv.org/content/10.1101/2025.01.24.634665v1`.

Zhenpeng Zhou, Steven Kearnes, Li Li, Richard N. Zare, and Patrick Riley. Optimization of molecules via deep reinforcement learning. *Scientific Reports*, 9:10752, 2019.

A Datasets

A.1 Pretraining Data Sources

We construct pretraining mixtures from a collection of publicly available molecular databases spanning distinct regions of chemical space (Figure 4):

- **ChEMBL**: bioactive molecules with experimentally measured activities, representing drug-like chemical space.
- **ZINC**: commercially available screening compounds.
- **COCONUT**: curated natural products spanning diverse biosynthetic classes.
- **HMDB**: human metabolites representing endogenous chemical space.

The UMAP projection in Figure 4 illustrates the relative positioning of these datasets in fingerprint space. The farthest-selected ZINC subset forms a partially separated cluster with reduced overlap with ChEMBL, confirming that the selection strategy expands structural diversity beyond drug-like regions.

All molecules are standardized (canonical SMILES, salt removal, and duplicate filtering) prior to mixture construction.

A.2 Fine-Tuning Dataset: SYK Inhibitors

For downstream adaptation, we use a curated dataset of SYK inhibitors with experimentally measured activities. Fine-tuning is performed on molecules with $pIC50 > 7$, focusing on highly active compounds to bias generation toward potent regions of the target space.

The SYK dataset occupies a localized region within the broader chemical landscape (Figure 4), partially overlapping with drug-like ChEMBL molecules but substantially smaller in coverage. This makes SYK an appropriate benchmark for evaluating whether diverse pretraining improves adaptability to a specific biological target.

A.3 Standardization and Filtering

All molecules are processed using RDKit. Invalid or unparsable SMILES strings are discarded. Canonical SMILES representations are generated for all valid molecules. Molecules exceeding the maximum SMILES length are removed. Duplicate molecules are eliminated based on canonical SMILES identity.

B Additional Experimental Details

B.1 Pretraining Configuration

Our molecular generator is based on a Transformer architecture optimized for SMILES sequence modeling. Table 1 summarizes the complete architecture and training configuration.

All models were trained on NVIDIA A6000 GPUs (512 GB memory). Pretraining required approximately 20 minutes for 500K molecules, while fine-tuning completed within 5-10 minutes per target.

B.2 Fine-Tuning Dynamics

Figure 5 shows score trajectories across fine-tuning epochs for all evaluated pretraining mixtures. All configurations exhibit a consistent pattern: scores rise sharply in the first 2–3 epochs, reach a peak, and then monotonically decline.

The primary cause of post-peak degradation is the collapse of scaffold diversity. As fine-tuning progresses, the model increasingly specializes toward the SYK inhibitor distribution, at the cost of

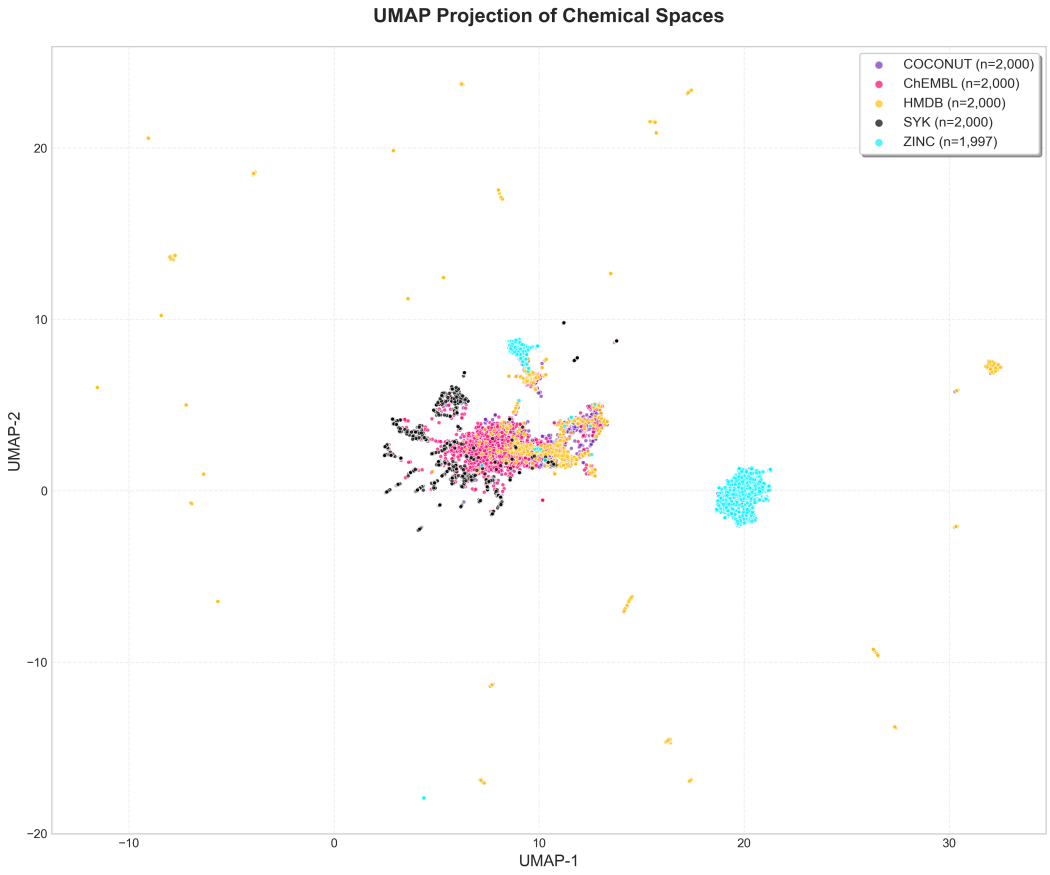


Figure 4: UMAP projection of molecular fingerprint space for COCONUT, ChEMBL, HMDB, SYK inhibitors, and a diversity-enriched ZINC subset. The ZINC subset was selected using a farthest-from-ChEMBL strategy, maximizing Tanimoto distance to drug-like molecules. This results in expansion into structurally distinct regions of chemical space. The SYK fine-tuning dataset occupies a localized subregion partially overlapping with ChEMBL, highlighting the domain adaptation challenge.

generating structurally redundant molecules. This reduces scaffold entropy, which in turn lowers the composite score despite continued improvement in predicted pIC50.

The recommended early stopping point (epoch 3) captures the optimal trade-off between target relevance and structural diversity. Beyond this point, extended fine-tuning enters an overfitting regime in which diversity loss outpaces activity gain.

Notably, the ZINC farthest-from-ChEMBL pretraining mixture achieves the highest peak score (9.09 at epoch 2), suggesting that broader pretraining coverage of non-drug-like chemical space provides a more robust initialization for target-specific adaptation. In contrast, ChEMBL-only baselines peak at lower scores and exhibit faster degradation, consistent with reduced initial diversity.

C Extended Results

C.1 Full Evaluation Tables

Table 2 summarizes all evaluated pretraining mixtures and curation policies, reporting the composite *Score* and its components (mean top-50 predicted pIC_{50} and scaffold entropy), together with auxiliary quality metrics (`Scaff.`, `Valid.`). The `SYK reference` row is included for context only.

Table 1: Model architecture and training configuration.

Parameter	Value
Architecture	
Layers	4
Hidden dimension	256
Attention heads	4
Head dimension	64
FFN dimension	1024
Vocabulary size	512
Max sequence length	128 tokens
Total parameters	2.8 M
Pretraining	
Epochs	5
Batch size	512
Learning rate	3×10^{-4}
LR schedule	Cosine decay
Warmup ratio	5 %
Weight decay	0.1
Optimizer	AdamW (β_1=0.9, β_2=0.95)
Fine-tuning	
Epochs	10
Batch size	64
Learning rate	1×10^{-4}
LR schedule	Cosine decay
Early stopping patience	5 epochs

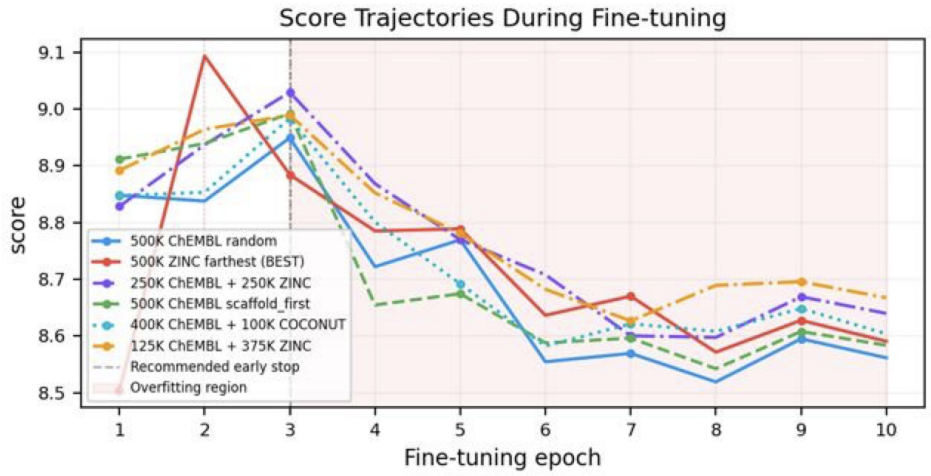


Figure 5: Score trajectories during fine-tuning across pretraining mixtures. Score is defined as scaffold entropy $\times$ mean top-50 predicted pIC50. All mixtures reach peak performance within epochs 2–3 (dashed vertical line), after which scores monotonically decline into the overfitting region (shaded). The ZINC farthest-from-ChEMBL mixture achieves the highest peak score (9.09 at epoch 2), while ChEMBL-only baselines peak lower and degrade faster.

Table 2: Extended Results Across Pretraining Mixtures. Score = scaffold entropy × mean top-50 predicted pIC50.

Mixture	Score	pIC50	Entropy	Scaff.	Valid.
ChEMBL (random)	8.949	9.482	0.944	27	0.666
ChEMBL (scaffold)	8.992	9.519	0.945	31	0.712
ChEMBL + ZINC (far.)	8.992	9.473	0.949	32	0.646
ZINC (farthest)	**9.093**	9.375	**0.970**	31	0.506
ChEMBL + ZINC (50/50)	9.030	9.569	0.944	30	0.689
125K ChEMBL + 375K ZINC	8.989	9.532	0.943	22	0.644
200K ChEMBL + 300K ZINC	8.637	9.610	0.899	19	0.752
300K ChEMBL + 200K ZINC	8.981	9.532	0.942	22	0.718
100K ChEMBL + 400K ZINC	9.012	9.343	0.965	30	0.617
150K ChEMBL + 350K ZINC	9.001	9.455	0.952	31	0.633
200K ChEMBL + 300K ZINC	9.021	9.455	0.954	31	0.633
400K ZINC + 100K COCONUT	8.990	9.455	0.951	31	0.633
ZINC (random)	8.945	9.288	0.963	28	0.502
COCONUT (scaffold)	8.729	9.498	0.919	23	0.483
COCONUT (random)	8.812	9.505	0.927	19	0.456
ZINC + COCONUT (400/100)	8.997	9.336	0.964	27	0.496
SYK reference	8.702	9.903	0.879	33	1.000

Table 3: Performance across five random seeds for baseline and best configuration. All metrics computed on top-1000 generated candidates after fine-tuning.

Model	**Seed**	**Score**	**pIC$_{50}$**	**Entropy**	**Validity**
Baseline	0	8.949	9.482	0.944	0.666
Baseline	1	8.929	9.533	0.937	0.674
Baseline	2	8.947	9.484	0.943	0.649
Baseline	3	8.983	9.539	0.942	0.669
Baseline	4	8.970	9.536	0.941	0.668
Exp_7 (ZINC farthest)	0	9.093	9.375	0.970	0.506
Exp_7 (ZINC farthest)	1	9.007	9.502	0.948	0.652
Exp_7 (ZINC farthest)	2	9.022	9.483	0.951	0.644
Exp_7 (ZINC farthest)	3	9.026	9.502	0.950	0.649
Exp_7 (ZINC farthest)	4	9.029	9.530	0.947	0.635

C.2 Statistical Validation Across Random Seeds

To ensure the robustness and reproducibility of our findings, we evaluated both the baseline configuration (500K random ChEMBL molecules) and the best-performing configuration (ZINC molecules maximally distant from ChEMBL) across five independent random seeds, controlling both dataset sampling and model weight initialization.

Table 3 presents the detailed results for each seed. Table 4 presents the detailed per-seed outcomes for each metric. To quantify whether the observed differences between the baseline and Exp_7 are statistically reliable across seeds, we additionally report a two-sided Welch's t-test for each metric in Table 4.

Table 4: Statistical significance of Baseline vs Exp_7 (ZINC farthest) across five random seeds ($n = 5$ per condition). Two-sided Welch's t-test. Values are reported as mean±sd; $\Delta = \mu_{\text{Exp_7}} - \mu_{\text{Base}}$. (After Bonferroni correction for 4 tests, $\alpha = 0.0125$, only Score remains significant.)

Metric	**Baseline (mean±sd)**	**Exp_7 (mean±sd)**	Δ	p
Score	8.9556±0.0211	9.0354±0.0333	+0.0798	**0.0026**
pIC$_{50}$	9.5148±0.0291	9.4784±0.0602	-0.0364	0.27
Entropy	0.9414±0.0027	0.9532±0.0095	+0.0118	**0.048**
Validity	0.6652±0.0095	0.6172±0.0625	-0.0480	0.16