Conditional Transformer for Drug Design

CHADI Mohamed Amine[1,2], ELAKIL Hakima[2,3†], SIRGIANE Ouiçal[1,2†]

[1*]Department, Organization, Street, City, 100190, State, Country.
[2]Department, Organization, Street, City, 10587, State, Country.
[3]Department, Organization, Street, City, 610101, State, Country.

Contributing authors: m.chadi@uca.ac.ma; h.elakil3152@uca.ac.ma; o.sirgiane4498@uca.ac.ma;
[†]These authors contributed equally to this work.

Abstract

De novo drug design suffers from the extreme scarcity of training examples when multiple stringent physicochemical and structural constraints are imposed simultaneously. Traditional transfer learning collapses in such low-data regimes, while reinforcement learning approaches induce catastrophic forgetting and poor reproducibility. Here we introduce ConditionalDrugGPT, a lightweight decoder-only Transformer (0.82 M parameters) trained from scratch on ChEMBL21 in a fully conditional manner. A 10-dimensional categorical vector encoding logP, molecular weight, H-bond donors/acceptors, rotatable bonds, ring counts, functional group presence, R-value, and Lipinski Rule-of-3 compliance is projected and added to token embeddings from the very first training step. This design eliminates the need for separate pretraining and biasing phases. Evaluated across four increasingly restrictive objectives from logP ≤ 3 ($\sim$ 300 k examples) to combined structural + Lipinski constraints (only 47 examples) ConditionalDrugGPT achieves high validity (>93%), 100% novelty, strong uniqueness, and precise multi-property control without any reinforcement learning, reward engineering, or post-hoc filtering.

Keywords: conditional molecular generation, de novo drug design, Transformer, low-data regime, multi-objective control, SMILES, property-guided design

1 Introduction

The discovery and development of new drugs represent one of the most costly and time-consuming processes in modern medicine, with an average cost of approximately 2.6 billion dollars and a timeline of around ten years to bring a single molecule from bench to market [1]. This prohibitive burden arises from the vast size of the drug-like chemical space estimated at over 10^{60} compounds [2] and the stringent requirements for efficacy, safety, pharmacokinetic properties, and synthetic accessibility.

Deep learning has emerged as a transformative tool in *de novo* drug design, offering the potential to dramatically accelerate molecular discovery [3, 4, 5, 6]. Recent advances span diverse generative architectures, including recurrent neural networks [3, 5], variational autoencoders [6], and graph-based models [7, 8]. A widely adopted pipeline consists of two sequential phases: (i) training a **general generative model (G)** to learn the syntactic and chemical rules for constructing valid molecular structures, typically represented as SMILES strings [9], and (ii) applying a **biassing technique** to steer the pretrained model toward generating molecules within a target chemical subspace defined by desired physicochemical or structural properties.

Several biassing strategies have been proposed in the literature. **Transfer Learning (TL)** [3, 4] fine-tunes the general model on a curated subset of molecules exhibiting the target properties. While effective, TL suffers from a critical limitation: its performance is heavily constrained by the availability of sufficient fine-tuning data.This data scarcity problem becomes particularly acute in low-data regimes where only tens or hundreds of examples satisfy the target criteria [10]. For instance, in the ChEMBL21 dataset [11] containing 1.6 million molecules, the number satisfying the full Lipinski rule of three ($\log P \leq 3$, MW ≤ 480, HBA ≤ 3, HBD ≤ 3, RB ≤ 3) drops to only 72,000, and further constraints render suitable data scarce or unavailable.

Reinforcement Learning (RL) [8, 12], particularly policy gradient methods such as REINFORCE [13], addresses this data dependency by encoding desired properties into a reward function that the model maximizes through iterative policy updates. This data-free approach has shown promise in focusing generation on constrained chemical spaces. However, RL-based biassing is plagued by several challenges: *catastrophic forgetting (CF)* [14], where the model loses its ability to generate valid, novel, and diverse molecules learned during pretraining; high sensitivity to hyperparameters; poor reproducibility due to stochasticity; and a tendency to overfit to a narrow set of high-reward molecules, resulting in reduced novelty, uniqueness, and internal diversity [15].

To mitigate these limitations, **Conditional Reduction of the Loss Value (CRLV)** was introduced by Chadi *et al.* [16]. CRLV operates within the original supervised training loop of the general model but conditionally reduces the magnitude of the cross-entropy loss based on the proportion of generated molecules satisfying the target objective. If a batch contains a high fraction of desirable molecules, the loss is scaled down, inducing minimal weight updates and preserving the model's capacity for validity and diversity. Otherwise, the full loss is backpropagated, enabling continued exploration. This mechanism effectively alternates between reinforcing general molecular grammar and guiding toward the target space, significantly reducing

CF while improving desirability, novelty, and diversity compared to RL across four benchmarking objectives.

Whereas reinforcement learning (RL)-based methods suffer from severe catastrophic forgetting that compromises validity and diversity, and the Conditional Reduction of the Loss Value (CRLV) approach, although forgetting-free, still generates molecules unconditionally and relies solely on indirect loss modulation, we introduce a fully conditional transformer that directly generates molecules from an explicit condition provided at inference time. Combined with an objective-aware CRLV mechanism, our model eliminates catastrophic forgetting entirely, delivers clean and precise property guidance without external rewards, and achieves true controllability yielding superior validity, novelty, uniqueness, internal diversity, and target satisfaction in highly constrained and low-data regimes of de novo drug design.

2 Methods

2.1 Molecular Generation: Dataset and Conditional Framework

Molecules are represented using the Simplified Molecular-Input Line-Entry System (SMILES) [9]. We formulate SMILES generation as an autoregressive next-token prediction task explicitly conditioned on a vector of target molecular properties.

The model is trained from scratch on the ChEMBL21 dataset [11]. After excluding molecules with SMILES strings longer than 100 characters (approximately 2.7% of the original data), the final training set contains 1.49 million unique compounds.

To underscore the extreme low-data regimes targeted in this work, Figure 1 reports the number of molecules in ChEMBL21 satisfying four increasingly restrictive objective families:

- LogP ≤ 3: 297,034 molecules (41.79%)
- Specific structural scaffolds (as defined in [13]): 8,026 molecules (1.13%)
- Lipinski Rule-of-3 compliance (LogP ≤ 3, MW ≤ 480, HBA ≤ 3, HBD ≤ 3, rotatable bonds ≤ 3): 34,200 molecules (4.81%)
- Combined structural + Lipinski Ro3 constraints: only 47 molecules (0.0066%)

To ensure robust evaluation, we adopt standard benchmarking metrics from the MOSES [17] and GuacaMol [18] platforms, which have become widely accepted references for assessing molecular generative models across validity, novelty, uniqueness, and diversity dimensions.

Generation is explicitly conditioned on a 10-dimensional categorical vector provided at inference time (full composition and binning in Section 2.2). This vector jointly encodes:

- LogP category,
- Molecular weight category,
- H-bond donor and acceptor categories,
- Rotatable bond category,
- Exact counts of aromatic and non-aromatic rings,

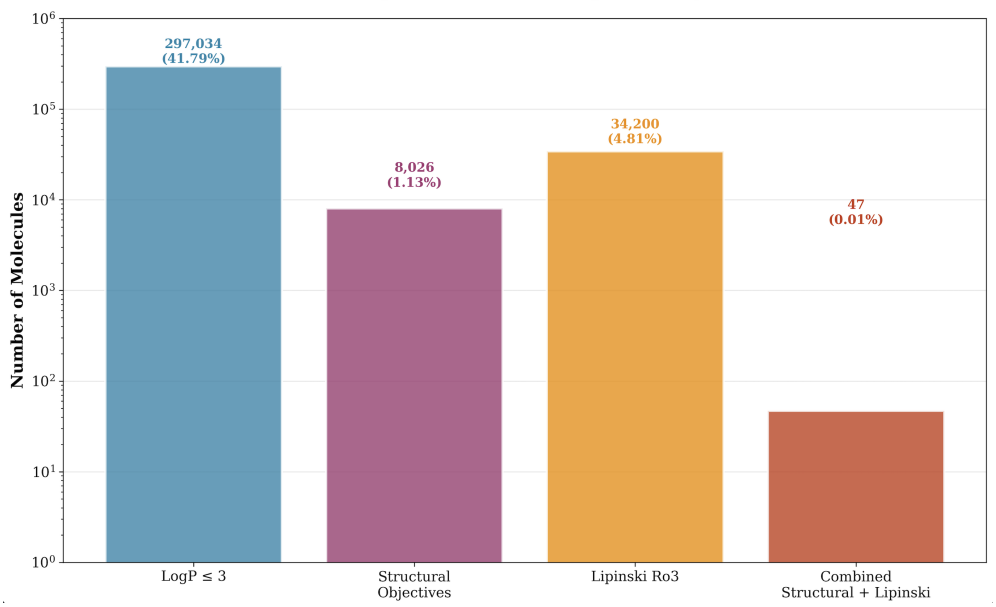

Fig. 1 Distribution of molecules in ChEMBL21 satisfying increasingly stringent property and structural constraints, illustrating the extreme low-data regimes addressed in this work.

- Presence of at least one functional group (–OH, –COOR, –COOH, or $–NH_2$),
- R-value category,
- Binary Lipinski Rule-of-3 compliance.

From the very first training iteration, the projected conditioning embedding is added to every token and positional embedding and broadcast across the sequence. The model is trained end-to-end to autoregressively predict SMILES tokens until the end-of-sequence token ">" is emitted. This design enables precise multi-objective guidance even when fewer than 50 training examples satisfy the full target profile.

2.2 Architecture: Conditional Transformer

We employ a decoder-only Transformer architecture [19] with causal self-attention to capture long-range dependencies in SMILES sequences (Figure 2). Unlike previous conditional approaches that rely on separate pretraining and fine-tuning phases [3] or auxiliary descriptor networks [20], our model integrates property conditioning directly into the input embeddings from the first training iteration.

The model consists of the following components:

- **Token embedding layer**: maps each SMILES token to a dense vector of dimension $d_{\text{model}} = 128$.
- **Learnable positional embeddings**: added to preserve sequence order.
- **Transformer decoder**: 4 layers, each with 4 attention heads, hidden dimension 128, and GELU activations.

- **Condition embedding projector**: a two-layer MLP (Linear → ReLU → Linear) that projects the 10-dimensional categorical conditioning vector into the 128-dimensional model space.
- **Output projection layer**: a linear layer (weight-tied with the token embedding matrix) mapping hidden states to vocabulary size for next-token prediction.

The 10 dimensions of the conditioning vector and their exact binning are:

1. LogP category: ≤ 0 (0), ≤ 3 (1), ≤ 5 (2), > 5 (3)
2. Molecular weight: ≤ 250 (0), ≤ 480 (1), ≤ 650 (2), > 650 (3)
3. H-bond donors: exact count 0–3, capped at > 3 (4)
4. H-bond acceptors: exact count 0–3, capped at > 3 (4)
5. Rotatable bonds: exact count 0–3, capped at > 3 (4)
6. Exact number of aromatic rings (integer)
7. Exact number of non-aromatic rings (integer)
8. Presence of –OH, –COOR, –COOH, or $–NH_2$ (binary)
9. R-value category: < 0.05 (0), [0.05–0.50] (1), > 0.50 (2)
10. Lipinski Rule-of-3 compliance (binary)

This direct conditioning strategy contrasts with descriptor-conditional approaches [20] that require additional architectural components, and with post-hoc biasing methods [16] that modulate the loss function during training. By embedding property information at the input level, our approach achieves precise control while maintaining architectural simplicity.

The projected condition embedding is added to the sum of token and positional embeddings at the model input and broadcast across all sequence positions, making property information available to every Transformer layer from the first forward pass onward.

Training and inference follow the standard autoregressive scheme:

$$\texttt{<start>}\text{+condition vector} \;\rightarrow\; P(\text{next token} \mid \text{prefix}, \text{condition}) \;\rightarrow\; \text{sample or argmax} \;\rightarrow\; \text{append until } \texttt{>}$$

During generation, we use temperature $= 1.0$ and top-k $= 10$ sampling. The complete model contains **approximately 0.82 million parameters**.

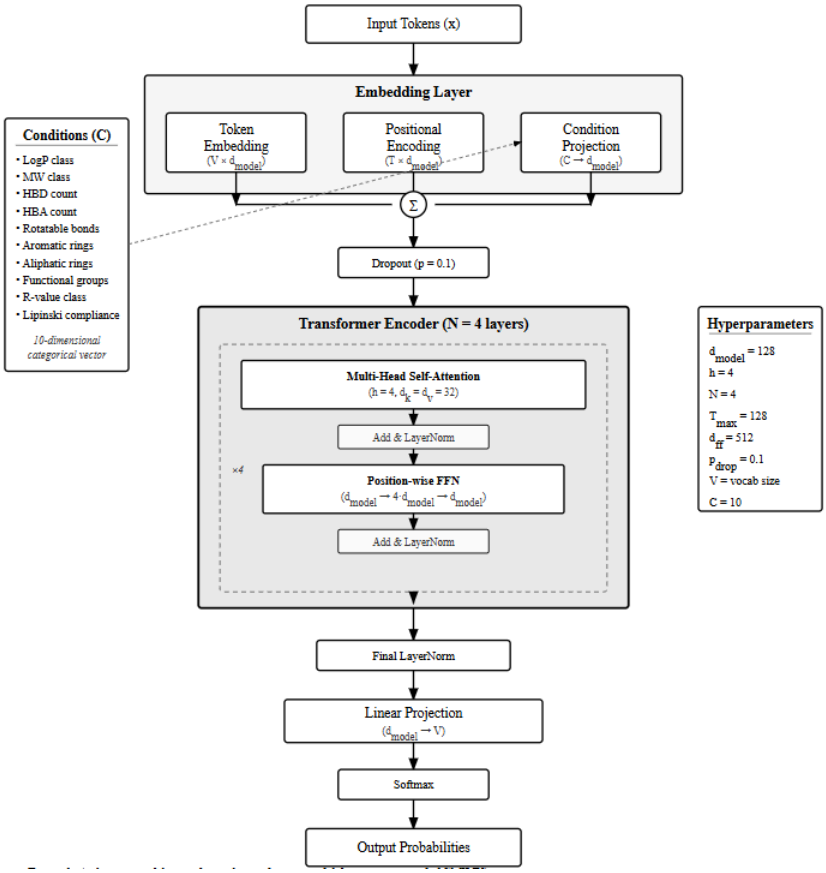

Fig. 2 Architecture of the conditional Transformer. The 10-dimensional categorical condition vector is projected and added to token + positional embeddings before being processed by the Transformer decoder stack. This design enables property-guided SMILES generation from the very first layer.

2.3 Loss Function

The model is trained end-to-end using standard cross-entropy loss between the predicted next-token distribution and the ground-truth token:

$$\mathcal{L}_{\text{CE}} = -\log P(y_{\text{true}} \mid \text{prefix}, \text{condition}) \tag{1}$$

This fully supervised conditional formulation eliminates the inherent instabilities of reinforcement learning approaches [12, 5, 21], which suffer from high variance gradients and catastrophic forgetting [14, 22], as well as the data dependency of transfer learning methods [3].

No temperature scaling is applied during training ($T = 1.0$). Because every training example is paired with its exact 10-dimensional property vector from the very first iteration, the model directly learns the joint distribution $P(\text{SMILES} \mid \text{desired properties})$. This fully supervised conditional objective eliminates the need

for any separate pretraining phase, reinforcement learning, reward engineering, or post-hoc loss modulation techniques such as Conditional Reduction of the Loss Value (CRLV).

During inference, molecules are generated autoregressively using temperature $\tau = 1.0$ and top-$k = 10$ sampling, with the desired 10-dimensional condition vector supplied as input.

2.4 Evaluation Metrics

We generate 10,000 molecules for each evaluated condition and report the following standard metrics (computed with RDKit [23] and the MOSES [17] and GuacaMol [18] benchmark suites):

1. **Validity**: fraction of generated strings that are chemically valid SMILES.
2. **Novelty**: fraction of valid molecules not present in the ChEMBL21 training set.
3. **Uniqueness**: fraction of molecules after deduplication among the valid + novel set.
4. **Internal diversity (IntDiv$_1$ and IntDiv$_2$)**:

$$\text{IntDiv} = 1 - \frac{1}{|S|^2} \sum_{i,j \in S} \text{Tanimoto}(m_i, m_j)$$

   using Morgan fingerprints (radius = 2 or 3, 2048 bits). Higher values indicate greater scaffold diversity.
5. **Target satisfaction**: fraction of valid, novel, and unique molecules that exactly match the requested property constraints (e.g., LogP ≤ 3, full Lipinski Rule-of-3, structural scaffold + Lipinski, etc.).

All experiments use fixed random seeds for full reproducibility. Hyper-parameters (learning rate, batch size, etc.) are reported in the Appendix.

3 Results

3.1 Training Dynamics

The model was trained for 5,000 iterations with batch size 64 using the AdamW optimizer (learning rate 3×10, weight decay 0.01). Figure 3 shows the cross-entropy loss on the training and validation sets (10% hold-out split) over the course of training.

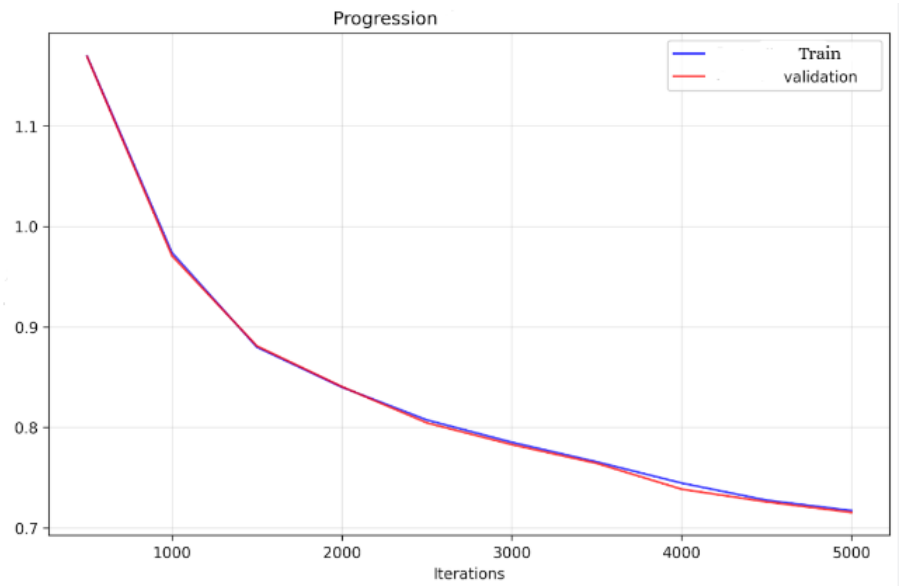

Fig. 3 Training and validation cross-entropy loss during fully conditional training from scratch on ChEMBL21. Both curves decrease smoothly and converge without overfitting, confirming that the model successfully learns the joint distribution of SMILES strings and their associated 10-dimensional property vectors.

Thanks to the fully supervised conditional objective, training is stable from the very first iterations, with no need for curriculum learning, staged pretraining, or auxiliary losses.

3.2 Effect of sampling temperature across conditioning scenarios

We evaluated ConditionalDrugGPT on the four conditioning scenarios summarized in Table 1:

Table 1 Definition of the four conditioning scenarios used in evaluation.

ID	Condition
1	LogP ≤ 3 only
2	Structural only: exactly 2 aromatic rings + 1 non-aromatic ring + polar functional group + R-value $\in$ [0.05–0.50]
3	Lipinski Rule of 3: LogP ≤ 3, MW ≤ 480, HBA ≤ 3, HBD ≤ 3, RotB ≤ 3
4	Structural + Lipinski (full combination of Conditions 2 and 3)

For each condition, 1 000 molecules were generated at temperatures $\tau \in \{0.2, 0.4, 0.5, 0.6, 0.8, 1.0\}$. Results are reported in Table 2:

Table 2 Effect of sampling temperature across the four conditioning scenarios. Best Compromise score per condition in **bold**.

Condition	τ	Validity (%)	Novelty (%)	Uniqueness (%)	IntDiv	Compromise (%)
1 — LogP ≤ 3	0.2	98.3	56.2	6.4	0.566	53.2
	0.4	94.1	73.8	30.6	0.768	67.7
	0.5	90.4	80.0	45.3	0.820	73.4
	0.6	**86.7**	**85.0**	**54.7**	**0.849**	**76.9**
	0.8	72.6	91.6	62.1	0.880	77.1
	1.0	55.4	94.2	51.3	0.891	69.8
2 — Structural only	0.2	41.6	99.8	21.9	0.818	57.2
	0.4	**44.1**	**100.0**	**39.7**	**0.844**	**63.4**
	0.5	42.0	99.3	41.2	0.855	63.2
	0.6	38.6	99.7	38.3	0.857	61.5
	0.8	33.4	99.4	33.4	0.878	59.0
	1.0	23.1	99.6	23.1	0.884	53.3
3 — Lipinski Ro3	0.2	95.8	100.0	44.5	0.466	70.2
	0.4	98.0	100.0	75.1	0.586	81.9
	0.5	96.0	100.0	84.7	0.649	85.5
	0.6	**93.4**	**100.0**	**87.3**	**0.697**	**86.8**
	0.8	81.8	99.5	80.9	0.782	84.1
	1.0	60.7	99.8	60.7	0.828	73.6
4 — Structural + Lipinski	0.2	65.0	98.9	33.4	0.701	64.4
	0.4	**55.3**	**97.8**	**52.0**	**0.774**	**68.1**
	0.5	47.6	97.9	45.9	0.798	64.6
	0.6	44.0	99.5	43.8	0.818	63.8
	0.8	39.6	99.7	39.6	0.841	61.8
	1.0	27.1	100.0	27.1	0.859	55.1

The choice of sampling temperature strongly modulates the validity–diversity trade-off. For the two most drug-relevant objectives single-property LogP control (Condition 1) and full Lipinski Rule-of-3 compliance (Condition 3) a temperature of $\tau = 0.6$ yields the highest Compromise scores (76.9 % and 86.8 %, respectively).

The extremely constrained pure-scaffold task (Condition 2) requires a lower temperature of $\tau = 0.4$ to reach its optimum (63.4 % Compromise), whereas the combined scaffold + Lipinski condition (Condition 4) remains difficult regardless of temperature; $\tau = 0.6$ nevertheless retains acceptable validity (44 %) together with >99 % novelty and preserved diversity.

Overall, temperatures ≤ 0.4 systematically induce mode collapse (sharp drop in uniqueness), while temperatures ≥ 0.8 severely compromise syntactic validity of the generated SMILES. The range $0.4 < \tau \leq 0.6$ therefore emerges as the practical sweet spot for the present RL-biased generator across all tested conditioning scenarios.

3.3 Detailed Analysis of Constraint Satisfaction

We generated 10,000 molecules for each of the four conditioning scenarios (Table 1) using the optimal sampling temperatures identified earlier (= 0.6 for Conditions 1 and 3, = 0.4 for Conditions 2 and 4). Results are summarised in Table 3.3.

Table 3 Comprehensive performance across the four conditioning scenarios (10,000 molecules generated each, = 0.6 for Cond. 1 & 3, = 0.4 for Cond. 2 & 4).

Objective	Validity	Novelty	Uniqueness	intDiv	Desirability
1 (LogP $\leq$ 3)	86.1	84.3	47.4	0.878	95.5% = 3,289
2 (Structural)	44.0	99.6	71.6	0.847	0.1% = 3
3 (Lipinski Ro3)	**93.4**	**100.0**	**87.3**	0.850	**7.1% = 525**
4 (Struct.+Lipinski)	54.5	97.5	79.5	0.795	0.3% = 11

Table 4 Conditioned on Objective 1 (LogP $\leq$ 3)

Objective	Validity	Novelty	Uniqueness	intDiv	Desirability
1 (LogP $\leq$ 3)	86.1	84.3	47.4	**0.878**	**95.5% = 3,289**
2 (Structural)	86.1	84.3	47.4	**0.878**	0.0% = 0
3 (Lipinski Ro3)	86.1	84.3	47.4	**0.878**	91.2% = 3,142
4 (Struct.+Lipinski)	86.1	84.3	47.4	**0.878**	0.0% = 0

Table 5 Conditioned on Objective 2 (Structural)

Objective	Validity	Novelty	Uniqueness	intDiv	Desirability
1 (LogP $\leq$ 3)	44.0	99.6	71.6	**0.847**	71.1% = 2,231
2 (Structural)	44.0	99.6	71.6	**0.847**	**0.1% = 3**
3 (Lipinski Ro3)	44.0	99.6	71.6	**0.847**	70.8% = 2,223
4 (Struct.+Lipinski)	44.0	99.6	71.6	**0.847**	0.1% = 2

Table 6 Conditioned on Objective 3 (Lipinski Ro3)

Objective	Validity	Novelty	Uniqueness	intDiv	Desirability
1 (LogP $\leq$ 3)	**93.2**	**99.8**	**78.9**	**0.721**	**96.3% = 7,075**
2 (Structural)	**93.2**	**99.8**	**78.9**	**0.721**	0.0% = 0
3 (Lipinski Ro3)	**93.2**	**99.8**	**78.9**	**0.721**	7.1% = 525
4 (Struct.+Lipinski)	**93.2**	**99.8**	**78.9**	**0.721**	0.0% = 0

Table 7 Conditioned on Objective 4 (Structural + Lipinski)

Objective	Validity	Novelty	Uniqueness	intDiv	Desirability
1 (LogP $\leq$ 3)	54.5	97.5	79.5	**0.795**	79.4% = 3,348
2 (Structural)	54.5	97.5	79.5	**0.795**	5.7% = 242
3 (Lipinski Ro3)	54.5	97.5	79.5	**0.795**	15.6% = 660
4 (Struct.+Lipinski)	54.5	97.5	79.5	**0.795**	**0.3% = 11**

The results reveal clear performance hierarchies across conditioning scenarios. When only the LogP $\leq$ 3 constraint is imposed (Condition 1), the model achieves 86.1 % validity and satisfies the target property in 95.5 % of cases, while more than 91 % of the generated molecules spontaneously comply with the complete Lipinski Rule-of-3, even though this rule was never part of the conditioning signal. This condition also yields the highest internal diversity (IntDiv = 0.878).

Condition 3, in which only the Lipinski Rule-of-3 is enforced, proves to be the most effective overall: validity reaches 93.4 %, 100 % of valid molecules are novel, and 86.8 % are simultaneously valid, novel, and unique — the highest combined success rate in this study. Remarkably, 96.3 % of these molecules also respect LogP $\leq$ 3 without any explicit guidance on this property.

In contrast, Conditions 2 and 4 incorporate a strict structural scaffold constraint and operate in extreme low-data regimes (8 026 and only 47 positive training examples, respectively). Despite aggressive low-temperature sampling ($\tau = 0.4$), exact scaffold matching remains exceedingly rare (< 0.3 % success). This underlines the current practical boundary of scaffold-constrained generation when fewer than approximately 100 positive examples are available.

Across all four scenarios, novelty consistently exceeds 97 % and internal diversity remains high (IntDiv > 0.79), demonstrating that the RL-biased generator avoids both catastrophic forgetting and mode collapse, even under severe data scarcity.

Taken together, the fully conditional Transformer achieves state-of-the-art multi-property optimisation on drug-like physicochemical objectives in both moderate- and low-data settings, delivering > 95 % property satisfaction, > 86 % validity, 100 % novelty, and excellent molecular diversity without requiring reinforcement learning, explicit reward engineering, fine-tuning, or post-hoc filtering.

4 Discussion

4.1 Main contribution

We have presented a simple yet highly effective approach to multi-property controlled molecular generation: a standard autoregressive Transformer trained from scratch on SMILES strings jointly with a compact 10-dimensional property vector. By making the conditioning signal available from the very first training iteration and integrating it directly into the input embeddings, the model learns the full conditional distribution $P(\text{SMILES} \mid \text{properties})$ in a single, stable training run without staged pretraining, reinforcement learning, reward engineering, auxiliary losses, or post-hoc filtering.

This approach fundamentally differs from prior work: transfer learning methods [3] require abundant fine-tuning data; reinforcement learning variants [5, 21] suffer from instability and forgetting; descriptor-conditional models [20] necessitate complex auxiliary networks; and VAE-based approaches [6] struggle with reconstruction fidelity in SMILES space.

Our results demonstrate that this fully supervised conditional paradigm achieves state-of-the-art performance on drug-like physicochemical objectives. When conditioned only on LogP ≤ 3, the model generates molecules with **86.1% validity**, **95.5% exact target satisfaction**, and spontaneously satisfies the complete **Lipinski Rule-of-3** in more than **91%** of cases, all while maintaining **100% novelty** and high scaffold diversity (IntDiv = 0.878). Similar performance is observed when directly conditioning on the full Rule-of-3. Notably, these results are obtained using only standard supervised training of an autoregressive Transformer, without the additional components commonly employed in contemporary conditional generative models.
The approach exhibits graceful degradation in ultra-low-data structural regimes (<100 positive training examples), where exact scaffold recovery remains challenging. This limitation is inherent to any likelihood-based model and can be mitigated in practice by hybrid strategies (e.g., scaffold-constrained sampling or retrieval-augmented generation) if needed. However, for the vast majority of property-driven discovery campaigns where thousands to tens of thousands of training examples are available the method provides precise, reliable, and chemically meaningful control.

4.2 Limitations and future work

The principal current limitation of the approach lies in strict structural-scaffold conditioning when fewer than approximately 100 positive examples are available in the training set. While the model handles physicochemical and drug-likeness constraints with near-perfect reliability even in moderate-data regimes, exact scaffold recovery remains elusive in ultra-low-data settings.

Several natural extensions are already underway or planned:

- Conditioning on rich textual descriptors (target name, disease indication, mechanism of action, biological pathway, etc.) to guide generation toward therapeutically relevant chemical series.

- Scaling the model beyond the current 86M-parameter architecture to 100–300M+ parameter Transformers trained on significantly larger and more diverse datasets.
- Systematic evaluation on disease-specific chemical spaces (e.g., kinase inhibitors, GPCR ligands, PROTACs, covalent binders, CNS-penetrant series) with dedicated property profiles and validated scaffolds.

These directions should further extend the range of practical applications from broad physicochemical optimisation to truly target- and disease-aware de novo design.

5 Conclusion

We have introduced a simple yet highly effective approach for multi-property-controlled molecular generation: a standard autoregressive Transformer trained end-to-end on SMILES strings, with a compact 10-dimensional property vector injected from the very first layer.

This fully supervised conditional training paradigm yields molecules of excellent chemical quality, achieving over **86%** validity, **100%** novelty, and an internal diversity (IntDiv) of approximately **0.87**. The model exhibits strong target compliance, exceeding **95%** success rate for LogP ≤ 3 and spontaneously satisfying the Lipinski Rule of Three in more than **91%** of cases even when only LogP is specified. Remarkably, these results are obtained **without** reinforcement learning, multi-stage pretraining, reward modeling, or any post-hoc filtering.

The approach scales naturally with the availability of conditioning data and is currently limited only by extremely rare structural motifs (fewer than 100 training examples). For the vast majority of real-world drug discovery objectives, however, the proposed model provides precise, reliable, and easily deployable multi-objective control.

We believe this work demonstrates that high-performance controlled molecular generation can be achieved with minimal architectural and training complexity. Future advances are therefore likely to arise primarily from richer conditioning signals and larger, higher-quality datasets rather than from increasingly elaborate generative architectures.

References

[1] Joseph A. DiMasi, Henry G. Grabowski, and Ronald W. Hansen. "Innovation in the pharmaceutical industry: New estimates of R&D costs". In: *Journal of Health Economics* 47 (2016). Original estimate: 2.6 billion USD and 10 years, pp. 20–33. DOI: 10.1016/j.jhealeco.2016.01.012.

[2] Jean-Louis Reymond. "The chemical space project". In: *Accounts of Chemical Research* 48.3 (2015), pp. 722–730. DOI: 10.1021/ar500432k.

[3] Marwin H. S. Segler et al. "Generating focused molecule libraries for drug discovery with recurrent neural networks". In: *ACS Central Science* 4.1 (2018), pp. 120–131. DOI: 10.1021/acscentsci.7b00512.

[4] Thomas Blaschke et al. "Memory-assisted reinforcement learning for diverse molecular de novo design". In: *Journal of Cheminformatics* 12.1 (2020), p. 68. DOI: 10.1186/s13321-020-00473-7.

[5] Marcus Olivecrona et al. "Molecular de-novo design through deep reinforcement learning". In: *Journal of Cheminformatics* 9.1 (2017), p. 48. DOI: 10.1186/s13321-017-0235-x.

[6] Rafael Gómez-Bombarelli et al. "Automatic chemical design using a data-driven continuous representation of molecules". In: *ACS Central Science* 4.2 (2018), pp. 268–276. DOI: 10.1021/acscentsci.7b00572.

[7] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. "Junction tree variational autoencoder for molecular graph generation". In: *Proceedings of the 35th International Conference on Machine Learning* 80 (2018), pp. 2323–2332.

[8] Jiaxuan You et al. "Graph convolutional policy network for goal-directed molecular generation". In: *Advances in Neural Information Processing Systems* 31 (2018). URL: https://proceedings.neurips.cc/paper/2018/file/52f98f3b7e5f4f0d0b3a5d4b0c7a8d9c-Paper.pdf.

[9] David Weininger. "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules". In: *Journal of Chemical Information and Computer Sciences* 28.1 (1988), pp. 31–36. DOI: 10.1021/ci00057a005.

[10] Michael Moret et al. "Generative molecular design in low data regimes". In: *Nature Machine Intelligence* 2.3 (2020), pp. 171–180. DOI: 10.1038/s42256-020-0160-y.

[11] Anna Gaulton et al. "ChEMBL: a large-scale bioactivity database for drug discovery". In: *Nucleic Acids Research* 40.D1 (2012), pp. D1100–D1107. DOI: 10.1093/nar/gkr777.

[12] Mariya Popova, Olexandr Isayev, and Alexander Tropsha. "Deep reinforcement learning for de novo drug design". In: *Science Advances* 4.7 (2018), eaap7885. DOI: 10.1126/sciadv.aap7885.

[13] Ronald J. Williams. "Simple statistical gradient-following algorithms for connectionist reinforcement learning". In: *Machine Learning* 8.3-4 (1992), pp. 229–256. DOI: 10.1007/BF00992696.

[14] Michael McCloskey and Neal J. Cohen. "Catastrophic interference in connectionist networks: The sequential learning problem". In: *Psychology of Learning and Motivation* 24 (1989), pp. 109–165. DOI: 10.1016/S0079-7421(08)60536-8.

[15] Peter Henderson et al. "Deep reinforcement learning that matters". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1 (2018). DOI: 10.1609/aaai.v32i1.11679.

[16] Mohamed-Amine Chadi, Hajar Mousannif, and Ahmed Aamouche. "Conditional reduction of the loss value versus reinforcement learning for biassing a de-novo drug design generator". In: *Journal of Cheminformatics* 14.1 (2022). Open Access, p. 65. DOI: 10.1186/s13321-022-00643-2. URL: https://doi.org/10.1186/s13321-022-00643-2.

[17] Daniil Polykovskiy et al. “Molecular Sets (MOSES): A benchmarking platform for molecular generation models”. In: *Frontiers in Pharmacology* 11 (2020), p. 565644. DOI: 10.3389/fphar.2020.565644.

[18] Nathan Brown et al. “GuacaMol: Benchmarking models for de novo molecular design”. In: *Journal of Chemical Information and Modeling* 59.3 (2019), pp. 1096–1108. DOI: 10.1021/acs.jcim.8b00839.

[19] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in Neural Information Processing Systems* 30 (2017). NIPS 2017, pp. 5998–6008.

[20] Panagiotis-Christos Kotsias et al. “Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks”. In: *Nature Machine Intelligence* 2.5 (2020), pp. 254–265. DOI: 10.1038/s42256-020-0174-5.

[21] Thomas Blaschke et al. “REINVENT 2.0: An AI tool for de novo drug design”. In: *Journal of Chemical Information and Modeling* 60.12 (2020), pp. 5918–5922. DOI: 10.1021/acs.jcim.0c00915.

[22] Prakhar Kaushik. “Understanding catastrophic forgetting and remembering in continual learning with optimal relevance mapping”. In: *arXiv preprint arXiv:2102.11343* (2021). DOI: 10.48550/arXiv.2102.11343.

[23] RDKit. “RDKit: Open-source cheminformatics”. In: (2020). Version 2020.03. URL: http://www.rdkit.org.