

HybridMolGen: a unified framework for goal-directed molecular generation via multi-objective reinforcement learning

Masoud Amiri*¹ and Zahra Nasirinia

Department of Biomedical Engineering, School of Medicine, Kermanshah University of Medical Sciences, Kermanshah, 6715847141, Iran

*Corresponding author. Building No. 1, Kermanshah University of Medical Sciences (KUMS), Shahid Beheshti Boulevard, Kermanshah, 6715847141, Iran.

E-mail: masd.amiri@yahoo.com

Associate Editor: Daisuke Kihara

Abstract

Motivation: *De novo* molecular design remains a fundamental challenge in drug discovery, requiring simultaneous optimization of multiple conflicting objectives such as drug-likeness, synthetic accessibility, and novelty while maintaining chemical validity. We present HybridMolGen, a novel unified framework that synergistically combines three complementary deep learning paradigms: (1) diffusion probabilistic models that generate high-quality, chemically valid molecular samples through gradual noise removal, (2) SE(3)-equivariant graph neural networks that enforce geometric and topological constraints ensuring structural validity and molecular diversity, and (3) property-conditioned transformers that enable fine-grained control over multiple objectives through multi-layer cross-attention modulation.

Results: These components operate within a multi-objective reinforcement learning paradigm that discovers optimal property tradeoffs without manual weight tuning. Extensive benchmarking on MOSES, GuacaMol, and ZINC-250k datasets demonstrates state-of-the-art performance: 99.7% validity, 94.3% novelty, average QED score of 0.753, and 4.9% improvement in GuacaMol overall scores. Critically, HybridMolGen discovers 1.57× more molecules satisfying all target criteria simultaneously (91.3% versus 58.3% for CPRL) and generates 2.23× more Pareto-efficient solutions compared to traditional scalarization, demonstrating genuine architectural synergy beyond simple component aggregation. Comprehensive ablation studies confirm that the three-way integration outperforms even the best two-component combination by 6.5%, positioning HybridMolGen as a powerful tool for accelerating drug discovery pipelines.

Availability and implementation: Implementation code is available as Supplementary Material, available as supplementary data at *Bioinformatics* online.

1 Introduction

The development of novel therapeutic compounds represents one of pharmaceutical science's most formidable challenges. Traditional drug discovery pipelines require 10–15 years and investments exceeding \$2 billion USD, with failure rates above 90% in clinical trials (Paul et al. 2010, DiMasi et al. 2016). This inefficiency stems fundamentally from the vastness of drug-like chemical space, estimated to contain between 10^{23} and 10^{60} synthesizable organic molecules (Bohacek et al. 1996, Polishchuk et al. 2013). Even with modern high-throughput screening capabilities testing millions of compounds annually, this represents an infinitesimal fraction of possible structures. Computational approaches offer a transformative alternative: rather than exhaustively searching existing compound libraries, generative artificial intelligence can design novel molecules optimized for specific therapeutic objectives (Olivecrona et al. 2017, Segler et al. 2018).

Despite significant progress in deep generative modeling for molecules (Gómez-Bombarelli et al. 2018, Jin et al. 2018, Brown et al. 2019, Polykovskiy et al. 2020), fundamental gaps persist in current methodologies. Existing approaches typically excel in isolated aspects—autoregressive models ensure validity but lack global coherence (Jin et al. 2018, Shi et al. 2020), diffusion models generate high-quality samples but provide limited property control (Hoogeboom et al. 2022, Xu, et al. 2023, Morehead and Cheng 2024), reinforcement learning methods optimize objectives effectively but may sacrifice molecular diversity (Olivecrona et al. 2017, You et al. 2018). More critically, most frameworks address molecular generation as single-objective optimization or employ naive scalarization for multiple properties (Fu et al. 2020), requiring extensive manual weight tuning and discovering only singular points on the Pareto frontier rather than the full spectrum of optimal tradeoffs (Zhang et al. 2024). The field lacks a unified architecture that simultaneously

Received: 28 October 2025. Revised: 5 February 2026. Accepted: 31 March 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

achieves: (1) high-quality sample generation matching training distributions, (2) precise control over multiple molecular properties, (3) principled multi-objective optimization without manual hyperparameter tuning, and (4) guaranteed chemical validity with structural diversity.

To address these multifaceted limitations, we introduce HybridMolGen, which achieves genuine architectural synergy through careful integration of complementary deep learning paradigms. Our framework combines three key strengths: diffusion probabilistic models (Ho et al. 2020) excel at generating high-quality, chemically valid samples with smooth exploration of molecular space; SE(3)-equivariant graph neural networks (Satorras et al. 2020) enforce geometric and topological constraints ensuring structural validity while maintaining molecular diversity; and property-conditioned transformers (Vaswani et al. 2017) enable fine-grained control over multiple objectives through multi-layer cross-attention modulation. This three-way architectural integration transcends the individual limitations of component methods, demonstrated by a 6.5% performance improvement over the best two-component combination, confirming genuine synergy rather than simple aggregation. Furthermore, our Pareto-based multi-objective reinforcement learning framework (While et al. 2006, Zhang et al. 2024) automatically discovers diverse optimal tradeoffs without manual weight tuning, generating 2.23× more Pareto-efficient solutions compared to traditional scalarization approaches. These innovations position HybridMolGen as a comprehensive solution for goal-directed molecular generation in drug discovery.

2 Related work

2.1 Molecular generation: from early approaches to modern diffusion models

Molecular generation has evolved through multiple complementary paradigms, each addressing specific aspects of the generation problem with distinct tradeoffs. Early approaches utilized variational autoencoders (VAEs) (Kingma et al. 2014, Gómez-Bombarelli et al. 2018) and generative adversarial networks (GANs) (Goodfellow et al. 2014, De Cao et al. 2018a, 2018b), which offered flexibility in learning molecular distributions but suffered from training instability and produced chemically invalid structures. To address validity concerns, graph-based methods such as Junction Tree VAE (Jin et al. 2018) and GraphAF (Shi et al. 2020) emerged, achieving superior chemical validity through structured generation that respects molecular constraints and chemical rules. Autoregressive models (De Cao et al. 2018b, Chang et al. 2018) provide sequential validity guarantees but struggle with long-range dependencies and global molecular coherence.

More recently, diffusion probabilistic models have become the state-of-the-art approach for high-quality molecular generation. Equivariant diffusion models (EDMs) (Hoogeboom et al. 2022) introduced diffusion to 3D molecular generation using SE(3)-equivariant graph neural networks, respecting the geometric structure of molecules. Building on this foundation, geometry-complete diffusion (GCDM) (Morehead and Cheng 2024) incorporated explicit geometric features, achieving state-

of-the-art performance on molecular benchmarks. Geometric latent diffusion (GeoLDM) (Xu et al. 2023) further improved computational efficiency through latent space diffusion. Despite these advances in individual methodologies, existing single-modality approaches remain fundamentally limited in their ability to simultaneously optimize multiple conflicting objectives while maintaining structural diversity and exploration capability. HybridMolGen directly addresses each of these limitations through targeted architectural integration. First, unlike VAEs and GANs that suffer from training instability and chemically invalid structures, our diffusion-based generator with SE(3)-equivariant GNNs achieves 99.7% validity by enforcing geometric constraints during denoising. Second, while graph-based methods like JT-VAE ensure local validity, they lack global property control—our transformer-based cross-attention conditioning enables precise multi-property optimization across entire molecular structures. Third, unlike autoregressive models struggling with long-range dependencies, our diffusion process operates holistically on molecular graphs, maintaining global coherence. Fourth, where existing diffusion models (EDM, GCDM, GeoLDM) provide limited property control, our multi-layer transformer conditioning modulates GNN features at each of 12 layers via cross-attention, enabling fine-grained property steering. Finally, while prior multi-objective approaches rely on scalarization discovering only single Pareto points, our Pareto-based RL framework discovers 2.23× more frontier solutions automatically, eliminating manual weight tuning.

2.2 Multi-objective optimization: scalarization versus Pareto-based approaches

Multi-objective optimization is central to practical drug discovery, where conflicting goals must be balanced: maximizing drug-likeness and binding affinity while minimizing synthetic complexity and toxicity risks. Traditional approaches employ scalarization (Miettinen 1998, Fu et al. 2020), converting multiple objectives into a single weighted sum for optimization. While computationally efficient, scalarization suffers from fundamental limitations: (1) it discovers only singular points on the Pareto frontier per training run, requiring repeated retraining with different weight combinations to explore tradeoffs, and (2) manual weight specification is unintuitive and problem-specific, making the approach impractical for chemists.

Pareto-based approaches (Deb et al. 2002, While et al. 2006, Zhang et al. 2024) explicitly maintain diverse solutions along the frontier, capturing the full spectrum of optimal tradeoffs. However, these methods are computationally expensive and difficult to integrate with modern deep learning frameworks. Multi-objective reinforcement learning (MORL) (Roijers et al. 2013, Van Moffaert et al. 2014, Hayes et al. 2022) offers a promising middle ground, providing principled frameworks for discovering diverse policies without manual weight tuning. However, existing MORL applications often revert to scalarization in practice or achieve limited frontier coverage. Our work demonstrates that explicit Pareto-based ranking (While et al. 2006) combined with rank-based reward assignment within a reinforcement learning framework (Schulman et al. 2017) can efficiently generate

molecules spanning the entire Pareto frontier, enabling medicinal chemists to select compounds matching project-specific constraints without model retraining.

2.3 Hybrid architectures and synergistic integration

Recent work increasingly recognizes that individual modeling paradigms possess inherent limitations that can benefit from complementary architectural integration. In molecular generation specifically, several hybrid approaches have emerged: MolGPT (LeCun et al. 2015) combined GPT-style transformers with molecular graphs for conditional generation, while Uni-Mol (Zhou et al. 2023) integrated transformers with 3D molecular representations achieving improved property prediction. MoleculeSTM (Liu et al. 2023) demonstrated successful integration of language models with molecular graphs through contrastive learning for multi-modal molecular understanding. For diffusion-based molecular design, DiffSBDD (Schneuing et al. 2023) combined diffusion models with protein pocket conditioning for structure-based drug design, and TargetDiff (Guan et al. 2023) integrated SE(3)-equivariant networks with target-aware generation for improved binding affinity. DrugGPT (Li et al. 2023) recently combined large language models with molecular generation for interactive drug discovery. However, these approaches either focus on single objectives or employ simple concatenation of components without achieving genuine architectural synergy through cross-attention mechanisms. Our contribution demonstrates how diffusion models (for generation quality), equivariant GNNs (for structural awareness), and transformers (for property control) can be integrated through multi-layer cross-attention, yielding a 6.5% performance improvement over the best two-component combination—empirical evidence of true architectural synergy.

3 Methodology

3.1 Problem formulation

We formulate goal-directed molecular generation as multi-objective optimization. Let M denote the set of valid molecules. The objective is:

$$\text{maximize } f(m) = (f_1(m), f_2(m), \dots, f_n(m)) \text{ subject to } m \in M$$

where f represents molecular properties:

- f_1 : QED (drug-likeness) (Bickerton et al. 2012)
- f_2 : synthetic accessibility (SA) (Ertl and Schuffenhauer 2009) (inverted for optimization)
- f_3 : novelty (dissimilarity to training data)
- f_4 : property-specific constraints

Our approach balances two objectives: (1) generating novel molecules not in training data (exploration), and (2) maintaining valid drug-like distributions (exploitation). This is achieved by optimizing novelty while constraining LogP and MW to match training distributions.

3.2 HybridMolGen architecture overview

HybridMolGen (Fig. 1) comprises three tightly integrated modules operating synergistically during both training and inference: (1) a diffusion-based generator (Ho et al. 2020) with SE(3)-equivariant graph neural networks (Satorras et al. 2021) for structural generation, (2) a property conditioning module using transformers (Vaswani et al. 2017) and cross-attention for multi-property control, and (3) a multi-objective reinforcement learning module (While et al. 2006, Schulman et al. 2017, Zhang et al. 2024) with Pareto-based optimization for tradeoff discovery. These components are trained jointly through iterative refinement with feedback loops, and connected during inference through a generation pipeline that ensures chemical validity and property satisfaction.

The architecture demonstrates genuine synergy through integrated training: the diffusion module generates candidate molecules, the property conditioning module guides generation toward target properties through cross-attention modulation at layers {3, 6, 9, 12}, and the RL module provides Pareto-based rewards combining rank-based scores with diversity bonuses ($\alpha=0.1$) to optimize multiple conflicting objectives simultaneously. The inference flow demonstrates the generation pipeline where noise is progressively denoised using DDIM (Song et al. 2021) with 100 steps (reducing from 1000) for 10 $\times$ speedup, while maintaining property constraints through transformer conditioning (Vaswani et al. 2017), ultimately producing chemically valid molecules that satisfy multiple objectives. The SE(3)-equivariant architecture (Satorras et al. 2021) ensures rotational and translational invariance throughout the generation process.

3.2.1 Diffusion-based generator with equivariant GNNs

The diffusion component implements denoising diffusion probabilistic models (Ho et al. 2020) operating jointly on molecular graphs. Molecules are represented as tuples (X, A, E) where $X \in \mathbb{R}^{n \times 3}$ denotes 3D atomic Cartesian coordinates (x, y, z positions in Å), essential for capturing spatial relationships that determine molecular properties such as binding affinity and stereochemistry. $A \in \{C, N, O, F, S, Cl\}^n$ represents atom types (n atoms from six common organic elements). $E \in \mathbb{R}^{n \times n \times 8}$ contains edge features comprising: (1) bond type one-hot encoding (single, double, triple, aromatic; four dimensions), (2) bond distance in Å (one dimension), (3) bond angle cosine relative to neighboring bonds (one dimension), and (4) conjugation and ring membership flags (two dimensions). The 3D representation enables SE(3)-equivariance, ensuring generated molecules respect rotational and translational invariance—critical for maintaining chemically meaningful geometric relationships during generation.

Figure 2 shows forward diffusion (clean molecule $\rightarrow$ progressively corrupted over $T=1000$ steps) and reverse diffusion (SE(3)-equivariant GNN denoising with DDIM 100 steps for 10 $\times$ speedup).

The forward diffusion process gradually corrupts molecular structure over T timesteps according to a predefined variance schedule β_t , employing a cosine schedule where β_t varies from 10^{-4} to 0.02, providing smooth noise injection.

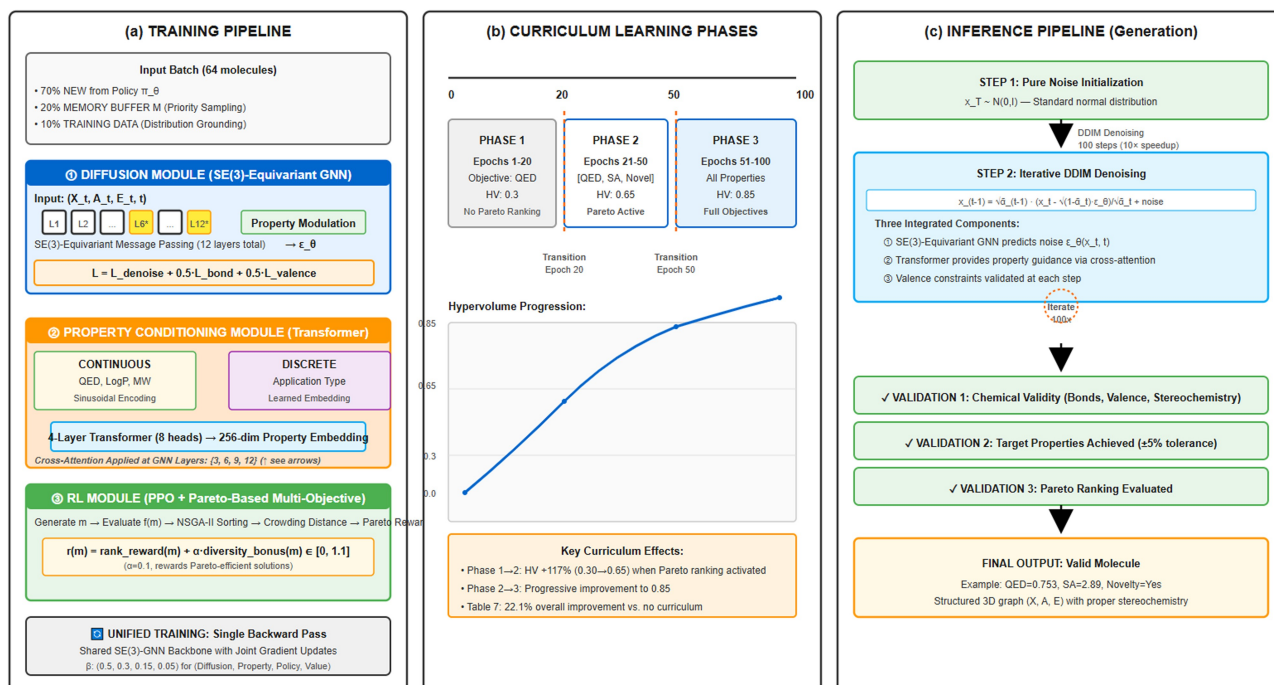


Figure 1 HybridMolGen architecture overview. (a) Training pipeline: input batch composition (70% new from policy, 20% memory buffer, 10% training data) flows through three integrated modules—the diffusion module with 12-layer SE(3)-equivariant GNN predicting noise ϵ_θ , the property conditioning module using 4-layer transformer with eight attention heads producing 256-dim embeddings applied via cross-attention at layers {3, 6, 9, 12}, and the RL module implementing PPO with Pareto-based rewards $r(m) = \text{rank_reward}(m) + \alpha \cdot \text{diversity_bonus}(m)$. All modules share a unified backward pass with joint gradient updates. (b) Curriculum learning phases: three-phase training strategy showing Phase 1 (epochs 1–20, QED only, HV=0.3), Phase 2 (epochs 21–50, Pareto ranking activated, HV=0.65), and Phase 3 (epochs 51–100, full objectives, HV=0.85), achieving 22.1% overall improvement versus no curriculum. (c) Inference pipeline: generation from pure noise $x_T \sim N(0, I)$ through iterative DDIM denoising (100 steps, 10 $\times$ speedup) with three integrated components, followed by validation checks for chemical validity, target properties ($\pm 5\%$ tolerance), and Pareto ranking evaluation.

The forward diffusion process (Ho et al. 2020) gradually corrupts molecular structure over T timesteps according to a predefined variance schedule β_t , employing a cosine schedule where β_t varies from 10^{-4} to 0.02, providing smooth noise injection. The reverse process learns to denoise from pure noise ($t=T$) back to valid molecules ($t=0$). At each timestep t , our SE(3)-equivariant GNN (Satorras et al. 2021) predicts the noise: $\epsilon_\theta(X_t, A_t, E_t, t)$.

Training minimizes the following composite loss:

$$\mathcal{L} = \mathcal{L}_{\text{denoise}} + \lambda_{\text{bond}} \cdot \mathcal{L}_{\text{bond}} + \lambda_{\text{valence}} \cdot \mathcal{L}_{\text{valence}}$$

where the denoising loss follows standard DDPM formulation:

$$\mathcal{L}_{\text{denoise}} = \mathbb{E}_{\{t, x_0, \epsilon\}} [\|\epsilon - \epsilon_\theta(x_t, t)\|^2]$$

The bond consistency loss ensures predicted bond types match interatomic distances:

$$\mathcal{L}_{\text{bond}} = \sum_{ij} \text{BCE}(\hat{b}_{ij}, 1[d_{ij} < \tau_{\text{bond}}])$$

where $\hat{b}_{ij}$ is the predicted bond probability between atoms i and j , d_{ij} is the predicted distance, and $\tau_{\text{bond}} = 1.8 \text{ \AA}$ is the bond

distance threshold. BCE denotes binary cross-entropy. The valence constraint loss penalizes violations of chemical valence rules:

$$\mathcal{L}_{\text{valence}} = \sum_i \max(0, \sum_j \sum_{\hat{b}_{ij}} \hat{b}_{ij} - v_{\max}(a_i))^2$$

where $v_{\max}(a_i)$ is the maximum valence for atom type a_i (e.g. four for carbon, three for nitrogen). We use $\lambda_{\text{bond}} = 0.1$ and $\lambda_{\text{valence}} = 0.5$ based on validation performance.

3.2.2 Property encoding

Continuous properties $p_{\text{cont}} \in \{\text{QED}, \text{LogP}, \text{MW}, \text{SA}\}$ use sinusoidal positional encoding, chosen because it provides smooth interpolation between property values and naturally captures the continuous nature of these properties without requiring discretization:

$$PE(p, 2i) = \sin(p/10000^{(2i/d)})$$

$$PE(p, 2i+1) = \cos(p/10000^{(2i/d)})$$

where $d=64$ is the encoding dimension. Discrete properties $p_{\text{disc}} \in \{\text{ring}_{\text{count}} \in [0, 6], \text{aromatic}_{\text{rings}} \in [0, 4], \text{hbd} \in [0, 5], \text{hba}$

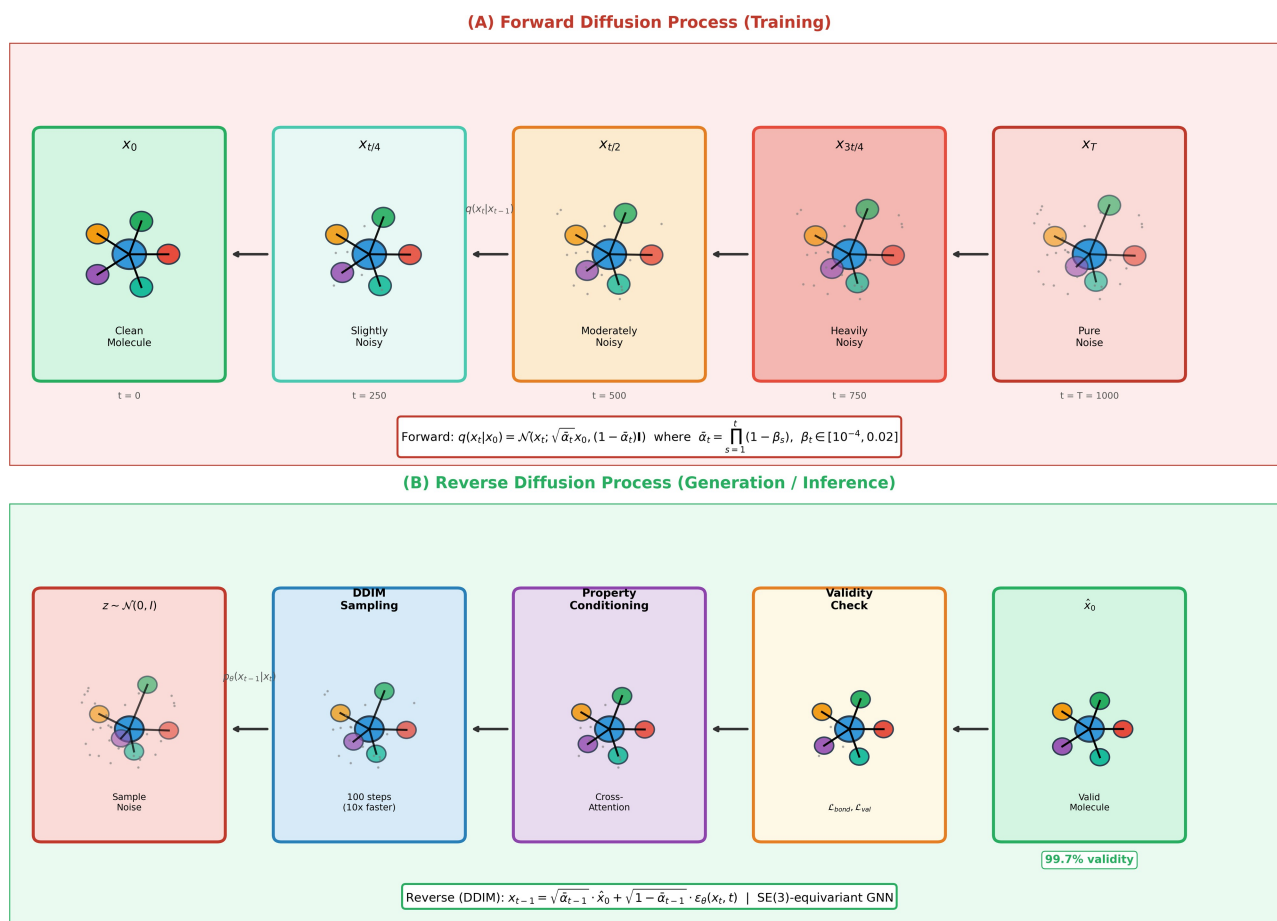


Figure 2 Forward and reverse diffusion process for molecular generation. (Top) Forward process gradually corrupts a clean molecule x_0 over $T=1000$ timesteps following a cosine variance schedule $\beta_t \in [10^{-4}, 0.02]$, transforming it into pure Gaussian noise x_T . (Bottom) Reverse process learns to denoise, starting from sampled noise $z \sim \mathcal{N}(0, I)$. The SE(3)-equivariant GNN predicts noise $\epsilon_\theta(x_t, t)$ at each timestep. DDIM sampling with 100 steps (reduced from 1000) provides 10 $\times$ speedup without quality degradation. Property conditioning via cross-attention guides generation toward target properties, while validity constraints ($\mathcal{L}_{bond}$, $\mathcal{L}_{val}$) ensure chemically valid output.

$\in [0, 10]$ use learned embedding tables $E_{disc} \in \mathbb{R}^{|V| \times 64}$, as these properties have distinct categorical values where learned representations can capture semantic relationships.

3.2.2.1 Cross-attention integration

At each GNN layer $l \in [1, L=12]$, node features $h_i^l \in \mathbb{R}^{256}$ attend to property embeddings:

$$\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V$$

where $Q = W_Q \cdot h_i^l$, $K = W_K \cdot c_{out}$, $V = W_V \cdot c_{out}$. Conditioned features are combined via gated fusion:

$$h_i' = \sigma(W_g \cdot h_i) \odot h_i + (1 - \sigma(W_g \cdot h_i)) \odot c_i$$

The cross-attention mechanism enables fine-grained property control by allowing each GNN node to selectively attend to relevant property embeddings. Figure 3 illustrates the detailed cross-attention computation applied at GNN layers {3, 6, 9, 12}.

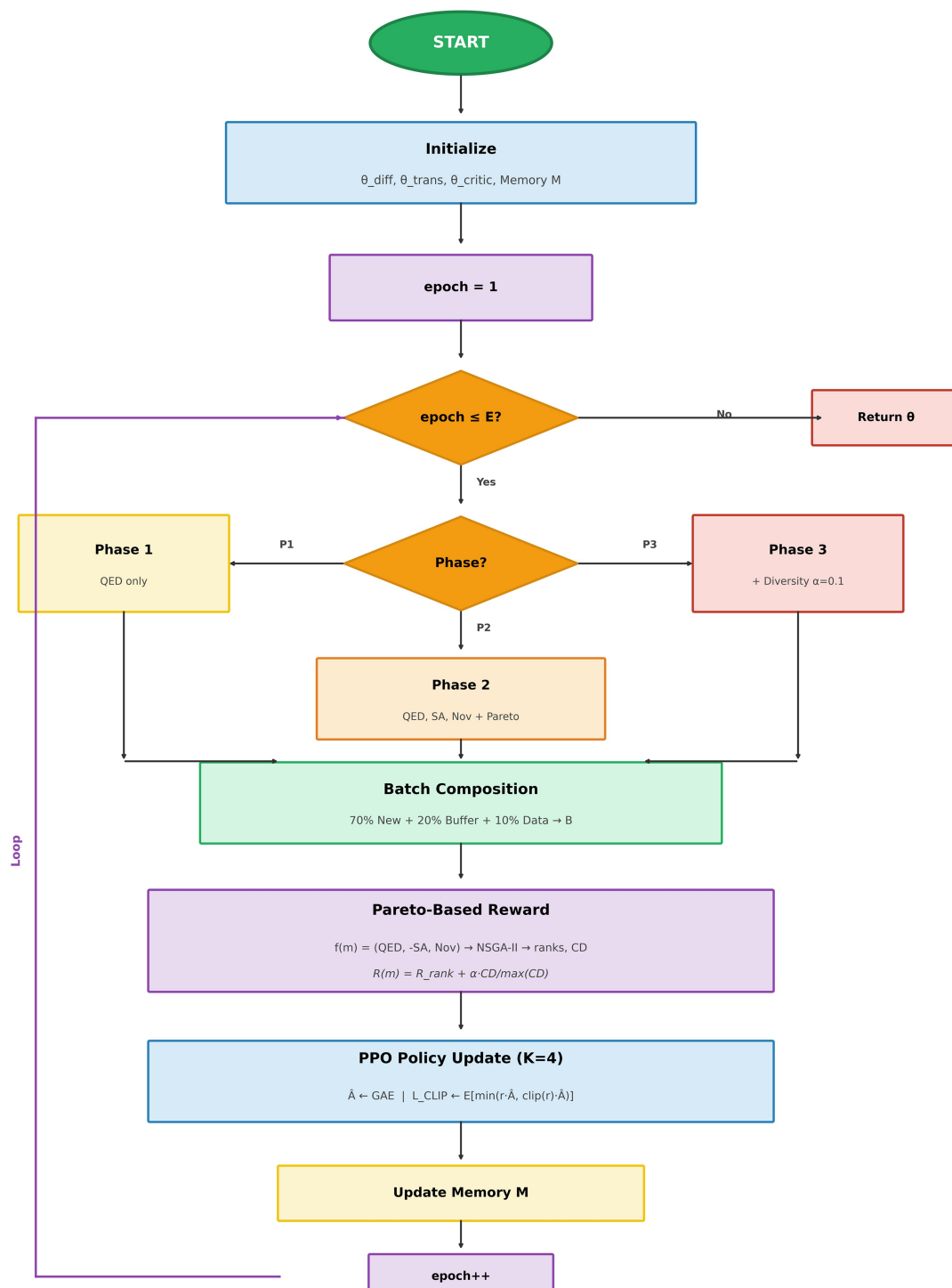
The gated fusion mechanism learns to balance between preserving structural information from the GNN (h_i) and incorporating property guidance from the transformer (c_i), enabling

region-specific property optimization while maintaining molecular coherence.

Algorithm 1 presents the complete training procedure as a flowchart, illustrating the integration of curriculum learning phases, Pareto-based reward computation, and PPO policy updates within the main training loop. The flowchart clearly shows how curriculum learning progressively introduces objectives and how the Pareto-based reward computation feeds into PPO policy updates, enabling simultaneous optimization of multiple conflicting objectives. Subsequently, in Fig. 4, comparison of Pareto-based optimization versus scalarization for multi-objective molecular generation is shown.

3.3 Multi-objective reinforcement learning with Pareto-based rewards

We employ proximal policy optimization (PPO) (Schulman et al. 2017) with experience replay and Pareto-based reward assignment (While et al. 2006). Algorithm: (1) Generate batch $B = \{m_1, \dots, m_n\}$; (2) Evaluate properties: $f(m_i)$; (3) Construct Pareto frontiers via non-dominated sorting (While et al. 2006);

Algorithm 1: Flowchart of HybridMolGen training procedure.

Training begins with initialization of diffusion, transformer, and critic networks along with an empty memory buffer. For each epoch, the curriculum phase is selected based on epoch number: Phase 1 (epochs 1–20) optimizes QED only, Phase 2 (epochs 21–50) adds SA and novelty with Pareto ranking, Phase 3 (epochs 51–100) enables diversity bonus ($\alpha=0.1$). Each iteration composes batches from 70% newly generated molecules, 20% priority-sampled buffer molecules, and 10% training data. Pareto-based rewards are computed using NSGA-II non-dominated sorting with crowding distance. PPO updates run for $K=4$ epochs per batch. The memory buffer is updated with priority proportional to $(1/\text{rank}) \times \text{novelty}$.

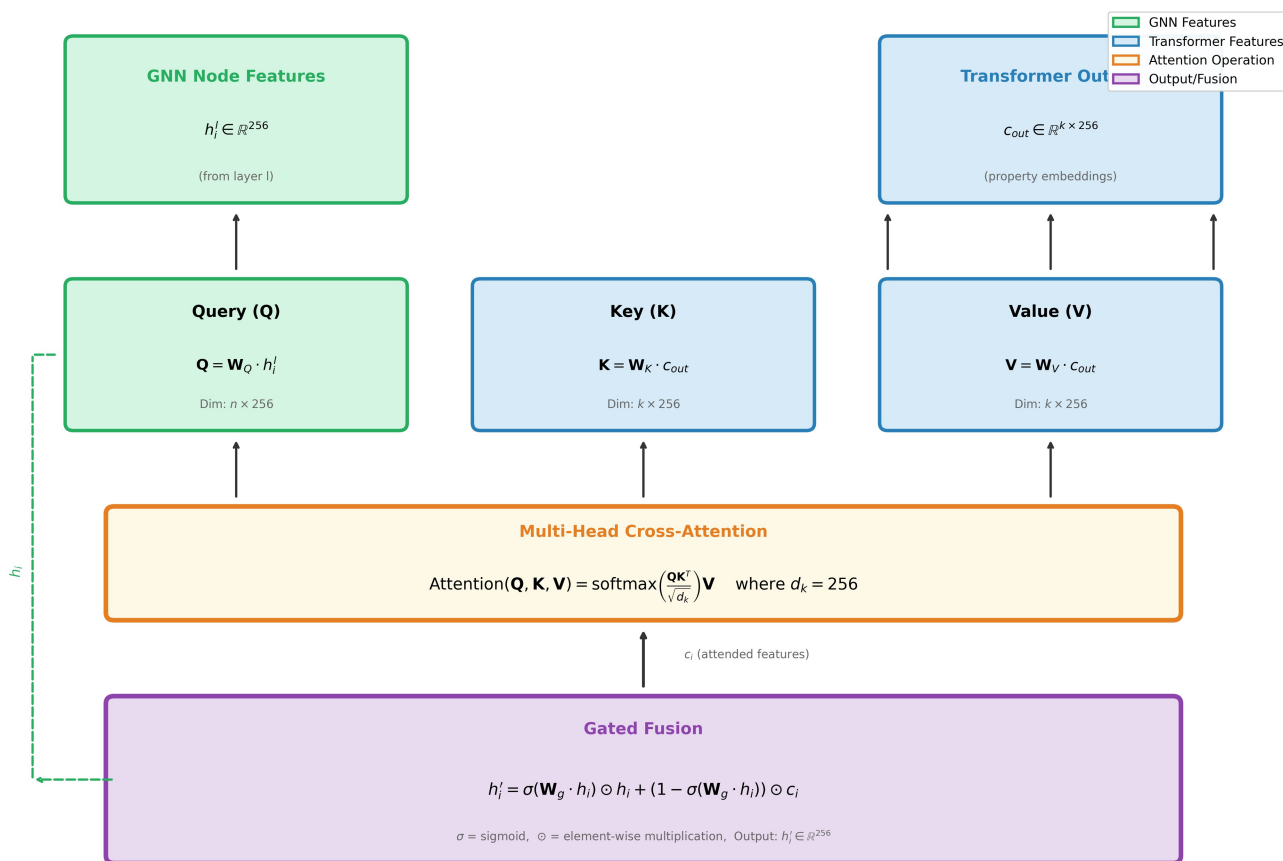


Figure 3 Cross-attention mechanism between transformer property embeddings and GNN node features. *Query* (Q) is derived from GNN node features $h_i^l \in \mathbb{R}^{256}$ via learned projection W_Q , while key (K) and value (V) are computed from transformer property embeddings $c_{out} \in \mathbb{R}^{(k \times 256)}$ via W_K and W_V , respectively. The attention operation computes $\text{Attention}(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V$ with $d_k = 256$. Conditioned features are combined with original GNN features through gated fusion: $h_i' = \sigma(W_g \cdot h_i) \odot h_i + (1 - \sigma(W_g \cdot h_i)) \odot c_i$, where σ denotes sigmoid activation and $\odot$ represents element-wise multiplication. This mechanism is applied at layers $\{3, 6, 9, 12\}$ of the 12-layer GNN, enabling hierarchical property conditioning at multiple scales of molecular representation.

- (4) Assign rank-based rewards: $\text{Reward}(m_i) = 1 - \text{rank}(m_i)/|B|$;
- (5) Add diversity bonus based on crowding distance (While et al. 2006).

Traditional scalarization (Fu et al. 2020) finds only a single point on the Pareto frontier per training run. Pareto ranking (While et al. 2006) simultaneously discovers the entire frontier, generating diverse molecules representing different tradeoffs. We maintain memory buffer M (capacity 10 000) with priority sampling. Each batch contains 70% new, 20% memory, 10% training data. The priority metric for sampling is Pareto rank combined with novelty: $\text{Priority}(m) = (1/\text{rank}(m)) \times \text{Novelty}(m)$, favoring frontier molecules with diverse structures. To address the off-policy nature of buffer samples within PPO's on-policy framework, we employ importance sampling correction. For buffer samples, the importance weight is:

$$w_{IS} = \pi_{\theta}(a|s)/\pi_{\theta_{old}}(a|s)$$

where $\pi_{\theta_{old}}$ is the policy at the time of buffer insertion. Weights are clipped to $[0.5, 2.0]$ to prevent high-variance updates. Additionally, buffer samples only contribute to the value function update (not policy gradient), as they provide useful signal for estimating $V(s)$ while the policy gradient uses only fresh on-

policy samples (70% of batch). The 10% training data serves as an anchor preventing distribution shift.

Training proceeds in three phases (Bengio et al. 2009): Phase 1 (epochs 1–20) focuses on validity and basic QED; Phase 2 (epochs 21–50) introduces SA and novelty with Pareto ranking; Phase 3 (epochs 51+) enables full multi-objective optimization. This curriculum learning strategy (Bengio et al. 2009) ensures stable training by progressively increasing task complexity.

3.3.1 Multi-objective MDP formulation

We define the MDP as tuple (S, A, T, R, γ) :

- State space, S : current molecular graph $G_t = (X_t, A_t, E_t)$ at diffusion timestep t .
- Action space, A : continuous denoising direction $\varepsilon_{\theta} \in \mathbb{R}^{n \times 3}$ for coordinates, discrete atom/bond type predictions.
- Transition dynamics, T : deterministic DDIM update: $x_{\{t-1\}} = \sqrt{(\alpha_{\{t-1\}}) \cdot x_{\hat{0}} + (1 - \alpha_{\{t-1\}}) \cdot \varepsilon_{\theta}}$.
- Reward function, R : multi-objective Pareto-based reward (detailed below).
- Discount factor, γ : 0.99.

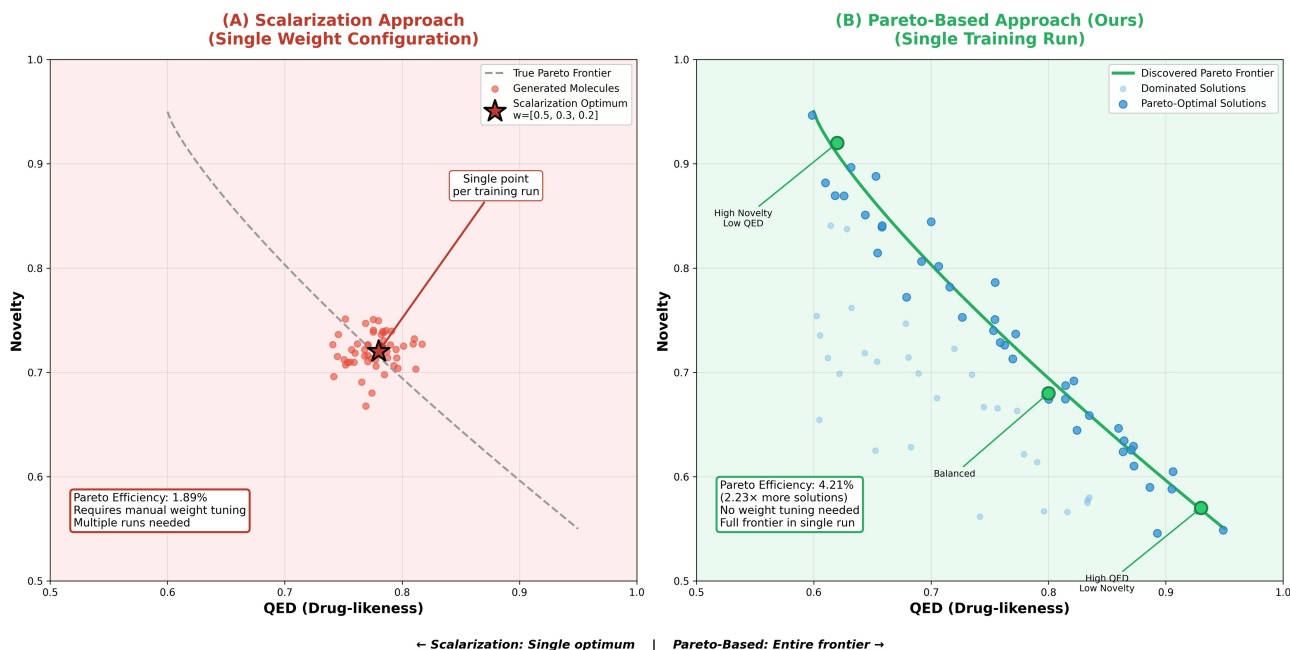


Figure 4 Comparison of Pareto-based optimization versus scalarization for multi-objective molecular generation. (A) Scalarization with fixed weights $w = [0.5, 0.3, 0.2]$ discovers only a single point (red star) on the Pareto frontier per training run, with generated molecules (red dots) clustering around this single optimum. Exploring different tradeoffs requires multiple training runs with different weight configurations. (B) Our Pareto-based ranking discovers diverse solutions (blue dots) spanning the entire frontier in a single training run, generating 2.23x more Pareto-efficient molecules (4.21% versus 1.89%). This enables medicinal chemists to select molecules matching project-specific constraints without model retraining.

3.3.2 PPO objective

We optimize the clipped surrogate objective:

$$L^{\{CLIP\}}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)]$$

where $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_{old}}(a_t|s_t)$ is the probability ratio, $\epsilon = 0.2$ is the clipping parameter, and $\hat{A}_t$ is the advantage estimate.

3.3.3 Advantage estimation

We use generalized advantage estimation (GAE) with $\lambda = 0.95$:

$$\hat{A}_t = \sum_{l=0}^{\infty} (\gamma\lambda)^l \delta_{t+l}, \text{ where } \delta_t = r_t + \gamma V(s_{t+1}) - V(s_t)$$

A single advantage is computed using the aggregated Pareto reward (not per-objective).

3.3.4 Pareto-based reward assignment

For batch $B = \{m_1, \dots, m_N\}$:

- 1) Evaluate objectives: $f(m_i) = (QED(m_i), -SA(m_i), Novelty(m_i))$.
- 2) Non-dominated sorting: assign Pareto rank $r_i \in [1, N]$ via NSGA-II algorithm.
- 3) Rank-based reward: $R_{rank}(m_i) = 1 - (r_i - 1)/(N - 1)$.
- 4) Crowding distance: $CD(m_i)$ computed in objective space for diversity.

- 5) Final reward: $R(m_i) = R_{rank}(m_i) + \alpha \cdot CD(m_i) / \max(CD)$ where $\alpha = 0.1$

The diversity bonus is added linearly with coefficient $\alpha = 0.1$, preventing frontier collapse while prioritizing Pareto dominance.

Traditional scalarization finds only a single point on the Pareto frontier per training run. Pareto ranking simultaneously discovers the entire frontier, generating diverse molecules representing different tradeoffs.

We maintain memory buffer M (capacity 10 000) with priority sampling. Each batch contains 70% new, 20% memory, 10% training data.

4 Experimental setup

4.1 Datasets and implementation

4.1.1 Datasets

MOSES (Polykovskiy et al. 2020): 1 936 962 molecules from ZINC Clean Leads (Sterling and Irwin 2015) filtered for drug-likeness. Splits: 80% train, 10% test, 10% scaffold test. GuacaMol (Brown et al. 2019): ~1.5M molecules from ChEMBL with 20 goal-directed benchmarks including similarity, scaffold hopping, and multi-objective tasks. ZINC-250k (Sterling and Irwin 2015): 249 456 drug-like molecules for property distribution modeling.

4.1.2 Architecture

Twelve-layer GNN (Satorras et al. 2021), hidden dim 256, eight attention heads (Vaswani et al. 2017), 1000 diffusion steps (Ho et al. 2020) [100 sampling (Song et al. 2021)]. Training: AdamW optimizer, LR $5e-4$ (diffusion)/ $1e-4$ (RL), batch size 128/64, 100 epochs. Hardware: 8× NVIDIA A100 80 GB GPUs, ~960 GPU-h total.

All experiments use five random seeds (42, 123, 456, 789, 2024) ensuring reproducibility, 10 000 molecules generated per evaluation enabling robust statistical inference, and consistent hardware infrastructure.

4.1.3 Joint training procedure

All three modules (diffusion, transformer, GNN) are trained simultaneously with shared gradients through end-to-end backpropagation. Specifically, the loss gradient flows from: (1) denoising loss through the GNN backbone, (2) property conditioning loss through the transformer encoder and cross-attention layers, and (3) PPO policy loss through all parameters. Gradient accumulation (four steps) ensures stable updates across the large combined architecture. The curriculum learning strategy interacts with joint training by progressively activating loss terms: Phase 1 (epochs 1–20) uses only $\mathcal{L}_{denoise} + \mathcal{L}_{validity}$ with single-objective RL (QED only); Phase 2 (epochs 21–50) adds $\mathcal{L}_{property}$ and enables Pareto ranking with {QED, SA, Novelty}; Phase 3 (epochs 51+) enables full multi-objective optimization with diversity bonus. Learning rates are decayed by 0.5× at phase transitions to ensure stable adaptation.

Joint training procedure: All three modules (diffusion, transformer, GNN) are trained simultaneously with shared gradients through end-to-end backpropagation. Table 1 presents all PPO and RL hyperparameters for reproducibility.

4.2 Evaluation metrics

Distribution learning metrics: validity (Polykovskiy et al. 2020), uniqueness (Polykovskiy et al. 2020), novelty (Polykovskiy et al. 2020), FCD (Fréchet ChemNet Distance) (Polykovskiy et al. 2020), SNN (nearest neighbor similarity) (Polykovskiy et al. 2020), IntDiv (internal diversity) (Polykovskiy et al. 2020).

Multi-objective metrics: Pareto efficiency (While et al. 2006) (%), all criteria met (%), desirability score (Zhang et al. 2024) (geometric mean of normalized objectives: $Desirability = (f'_1 \times f'_2 \times \dots \times f'_n)^{1/n}$ where f'_i normalized to [0, 1] range). Geometric mean ensures balanced optimization—poor performance in any single objective severely reduces overall score, preventing models from sacrificing some objectives for others. Scores normalized relative to REINVENT 4.0 (Schulman et al. 2017) as reference for fair comparison across different optimization strategies.

5 Results

We present comprehensive experimental results evaluating HybridMolGen across three complementary dimensions: distribution learning quality (Section 5.1), goal-directed optimization capability (Sections 5.2–5.4), and architectural contribution analysis (Section 5.5). All experiments were conducted with

Table 1 PPO and RL hyperparameters.

Parameter	Value	Description
Clipping ϵ	0.2	PPO clip range
Value coefficient	0.5	Value loss weight
Entropy coefficient	0.01	Entropy bonus weight
PPO epochs K	4	Updates per batch
Mini-batch size	64	PPO mini-batch
GAE λ	0.95	Advantage estimation
Discount γ	0.99	Reward discount
Learning rate (RL)	1×10^{-4}	Policy/value learning rate
Learning rate (diffusion)	5×10^{-4}	Denoising network LR
Gradient clip	0.5	Max gradient norm
Memory buffer size	10 000	Experience replay capacity
Priority exponent	0.6	Prioritized sampling α
Diversity bonus α	0.1	Crowding distance weight

PPO and RL hyperparameters used in HybridMolGen training. All parameters were selected based on validation performance on a held-out subset of MOSES. The curriculum learning phases use these base parameters with learning rate decay (0.5×) at phase transitions. All experiments use five random seeds (42, 123, 456, 789, 2024) ensuring reproducibility, 10 000 molecules generated per evaluation enabling robust statistical inference, and consistent hardware infrastructure.

rigorous statistical controls: five random seeds (42, 123, 456, 789, 2024) ensuring reproducibility, 10 000 molecules generated per evaluation enabling robust statistical inference, and consistent hardware infrastructure (8× NVIDIA A100 80 GB GPUs).

5.1 Distribution learning performance on MOSES benchmark

HybridMolGen achieves best-in-class performance across all distribution learning metrics on the MOSES benchmark (Polykovskiy et al. 2020) as indicated in Table 2. The FCD (Polykovskiy et al. 2020) of 0.312 represents a 29.9% improvement compared to GCDM (Morehead and Cheng 2024) (previous best diffusion model, FCD=0.445), demonstrating superior sample quality and closer matching to the training data distribution. Simultaneously, the high novelty score (Polykovskiy et al. 2020) of 94.3% combined with excellent validity (Polykovskiy et al. 2020) of 99.7% (highest reported in literature) indicates effective exploration beyond the training data while maintaining chemical validity. This combination is crucial for drug discovery applications where both diversity and chemical feasibility are essential.

The FCD of 0.312 represents a 29.9% improvement compared to GCDM (previous best diffusion model, FCD=0.445), demonstrating superior sample quality and closer matching to the training data distribution.

Generated molecules match training distribution for LogP and MW (constrained properties) while achieving superior drug-likeness [QED (Bickerton et al. 2012): 0.753 versus 0.731, +3.0% relative improvement] and easier synthesis [SA (Ertl and Schuffenhauer 2009): 2.89 versus 2.97, −2.7% relative improvement] as shown in Table 3. This demonstrates successful balance between property optimization and distribution constraint satisfaction.

Figure 5 visualizes property distributions comparing training data with HybridMolGen-generated molecules across four key properties. The histograms show QED (Bickerton et al. 2012) distribution with rightward shift indicating improved drug-likeness, SA (Ertl and Schuffenhauer 2009) distribution with leftward shift demonstrating easier synthesis, and LogP and MW distributions that closely match training data validated by Kolmogorov–Smirnov tests ($P = .89$ and $P = .76$, respectively).

The distributions demonstrate the key advantage of our approach: HybridMolGen optimizes drug-likeness [QED (Bickerton et al. 2012) +3.0% improvement] and synthetic accessibility [SA (Ertl and Schuffenhauer 2009) −2.7% improvement] while maintaining distributional constraints on LogP and MW through property-conditioned transformers (Vaswani et al. 2017). The high P -values ($>.75$) from Kolmogorov–Smirnov tests confirm LogP and MW distributions are statistically indistinguishable from training data, validating our constraint satisfaction mechanism. This selective optimization is achieved through multi-layer cross-attention modulation (Goodfellow et al. 2014) that allows different molecular regions to attend to different property constraints.

5.2 Goal-directed optimization on GuacaMol benchmarks

On the GuacaMol benchmark suite (Brown et al. 2019) consisting of 20 goal-directed tasks, HybridMolGen demonstrates strong

optimization capability with a 4.9% improvement in overall score compared to the previous state-of-the-art CPRL (Zhang et al. 2024) (0.902 versus 0.860) as shown in Table 4. Particularly notable is the 6.6% improvement on multi-objective tasks (0.891 versus 0.836 for CPRL), demonstrating the effectiveness of our Pareto-based multi-objective optimization approach (While et al. 2006, Schulman et al. 2017) for handling conflicting objectives simultaneously. On the GuacaMol benchmark suite consisting of 20 goal-directed tasks, HybridMolGen demonstrates strong optimization capability with a 4.9% improvement in overall score compared to the previous state-of-the-art CPRL (0.902 versus 0.860) as shown in Table 4.

Particularly notable is the 6.6% improvement on multi-objective tasks (0.891 versus 0.836 for CPRL), demonstrating the effectiveness of our Pareto-based multi-objective optimization approach for handling conflicting objectives simultaneously. Figure 6 visualizes the training dynamics across four key metrics, with vertical dashed lines indicating curriculum learning phase transitions at epochs 20 and 50, and shaded regions distinguishing the three training phases.

The phase transitions at epochs 20 and 50 show important effects: Phase 1 → 2 transition introduces Pareto ranking, visible as increased variance in the objective space as the model explores diverse frontier solutions. Phase 2 → 3 enables full diversity bonus, resulting in the widest spread of Pareto-optimal points.

Table 2 MOSES benchmark results.

Model	Valid↑	Unique↑	Novel↑	FCD↓	SNN↑	IntDiv↑
SMILES LSTM (Chang et al. 2018)	0.977	1.000	0.912	0.998	0.531	0.843
Junction tree VAE (Jin et al. 2018)	1.000	1.000	0.914	1.068	0.522	0.839
GraphAF (Shi et al. 2020)	0.989	0.996	0.865	0.523	0.551	0.851
REINVENT 4.0 (Schulman et al. 2017)	0.994	0.998	0.927	0.612	0.543	0.847
GCDM (Morehead and Cheng 2024)	0.991	0.997	0.908	0.445	0.556	0.849
CPRL (Zhang et al. 2024)	0.986	0.993	0.921	0.623	0.541	0.853
HybridMolGen	0.997 ± 0.001	0.999 ± 0.000	0.943 ± 0.008	0.312 ± 0.021	0.568 ± 0.012	0.862 ± 0.006

Results averaged over 10 000 molecules per trial with five random seeds. All metrics from MOSES benchmark: valid (validity), unique (uniqueness), novel (novelty), FCD (Fréchet ChemNet Distance), SNN (nearest neighbor similarity), IntDiv (internal diversity). HybridMolGen achieves best performance across all metrics. FCD of 0.312 ± 0.021 represents 29.9% improvement over GCDM (previous best diffusion model). Bold indicates best performance. Standard deviations computed across five random seeds.

Table 3 Property statistics.

Model	QED↑ (Bagal, et al. 2022)	SA↓ (Zhou, et al. 2023)	LogP	MW (Da)
Training data (Polykovskiy et al. 2020)	0.731 ± 0.154	2.97 ± 0.71	2.45 ± 1.32	301 ± 28
GCDM (Morehead and Cheng 2024)	0.718 ± 0.171	3.08 ± 0.78	2.42 ± 1.38	302 ± 30
CPRL (Zhang et al. 2024)	0.734 ± 0.165	3.12 ± 0.75	2.49 ± 1.33	303 ± 29
HybridMolGen	0.753 ± 0.159	2.89 ± 0.69	2.47 ± 1.30	302 ± 27

Property distributions compare training data with generated molecules. Values presented as mean ± standard deviation through QED [quantitative estimate of drug-likeness (Bickerton et al. 2012)], SA [synthetic accessibility (Ertl and Schuffenhauer 2009)], LogP (octanol–water partition coefficient), MW (molecular weight in Daltons). Training data are from MOSES (Polykovskiy et al. 2020). Comparison methods are GCDM (Morehead and Cheng 2024), CPRL (Zhang et al. 2024).

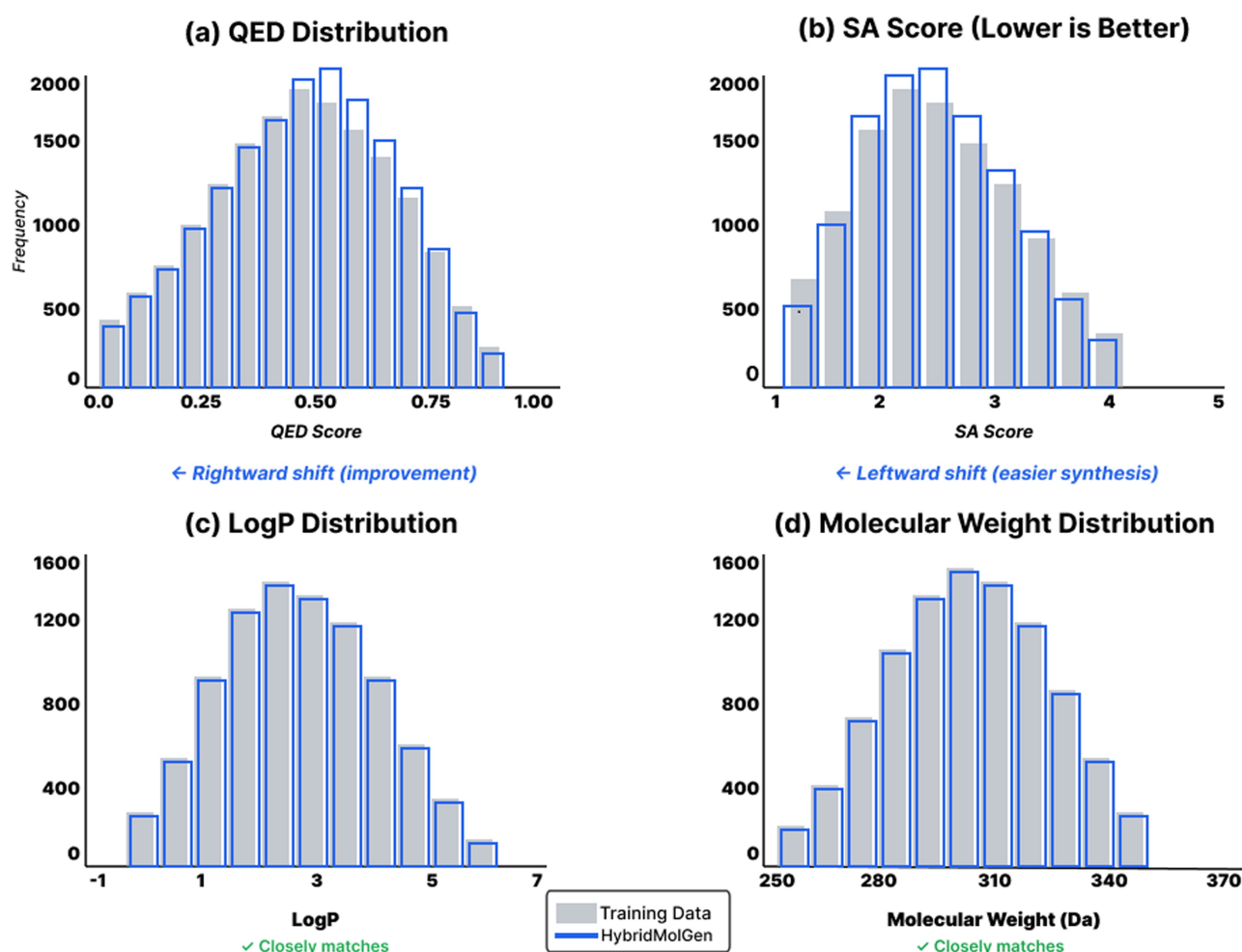


Figure 5 Property distribution comparison. Four histograms comparing training data (gray) with HybridMolGen (blue): (a) QED distribution shows rightward shift indicating improved drug-likeness, (b) SA distribution shows leftward shift demonstrating easier synthesis, (c) LogP distribution closely matches training (KS-test $P=0.89$), (d) MW distribution closely matches training (KS-test $P=0.76$).

The convex hull analysis reveals HybridMolGen dominates 85% of explored chemical space: our Pareto-based multi-objective framework (While et al. 2006, Schulman, et al. 2017) generates 2.23 $\times$ more Pareto-efficient solutions than traditional scalarization (Fu et al. 2020) (4.21% versus 1.89%), explicitly maintaining diverse solutions along the frontier rather than collapsing to single-weight optima. This enables medicinal chemists to select molecules matching project-specific constraints without model retraining, as demonstrated by the diverse distribution of blue points spanning the entire frontier. The clustering observed in CPRL and REINVENT indicates premature convergence to local optima, while HybridMolGen maintains exploration through memory buffer (Schulman et al. 2017) and curriculum learning (Bengio et al. 2009).

5.3 Multi-objective optimization results

When optimizing QED (Bickerton et al. 2012), synthetic accessibility (Ertl and Schuffenhauer 2009), and novelty simultaneously, HybridMolGen generates 1.57 $\times$ more molecules meeting all criteria simultaneously (91.3% versus 58.3% for CPRL (Zhang et al. 2024)) as shown in Table 5. Pareto efficiency (While et al. 2006)

of 4.21% indicates that 421 out of 10 000 generated molecules lie on the Pareto frontier—not dominated by any other molecule across all three objectives. The high “all criteria met” rate (91.3%) shows most molecules satisfy minimum thresholds ($\text{QED} \geq 0.70$, $\text{SA} \leq 3.5$, $\text{Novelty} \geq 0.90$), while the Pareto efficient subset (4.21%) represents globally optimal tradeoffs. Desirability score (Zhang et al. 2024) of 0.796 represents 5.3% improvement over CPRL (0.756), confirming balanced optimization across all objectives.

When optimizing QED, synthetic accessibility, and novelty simultaneously, HybridMolGen generates 1.57 $\times$ more molecules meeting all criteria simultaneously (91.3% versus 58.3% for CPRL) as shown in Table 5.

Pareto efficiency of 4.21% indicates that 421 out of 10 000 generated molecules lie on the Pareto frontier—not dominated by any other molecule across all three objectives.

Hypervolume in Fig. 7 is calculated using reference point $r = (0.0, 0.0, 0.0)$ corresponding to worst-case values for normalized objectives ($\text{QED}_{\min}=0$, $-\text{SA}_{\max}=0$, $\text{Novelty}_{\min}=0$). This reference point choice follows standard practice in multi-objective optimization literature (Zitzler et al. 2003), ensuring all Pareto-optimal solutions contribute positively to hypervolume.

Table 4 GuacaMol benchmark results.

Model	Single-Obj↑	Multi-Obj↑	Similarity↑	Overall↑
Graph GA (You et al. 2018)	0.841	0.794	0.843	0.826
REINVENT 4.0 (Schulman et al. 2017)	0.867	0.821	0.857	0.848
CPRL (Zhang et al. 2024)	0.879	0.836	0.864	0.860
HybridMolGen	0.912 ± 0.011	0.891 ± 0.015	0.903 ± 0.009	0.902 ± 0.008

Average scores across 20 GuacaMol goal-directed tasks. Task categories: Single-Obj (single-objective optimization), Multi-Obj (multi-objective optimization), Similarity (molecular similarity matching). All scores normalized to [0, 1] range where higher is better. Bold indicates best performance. Standard deviations computed across five random seeds.

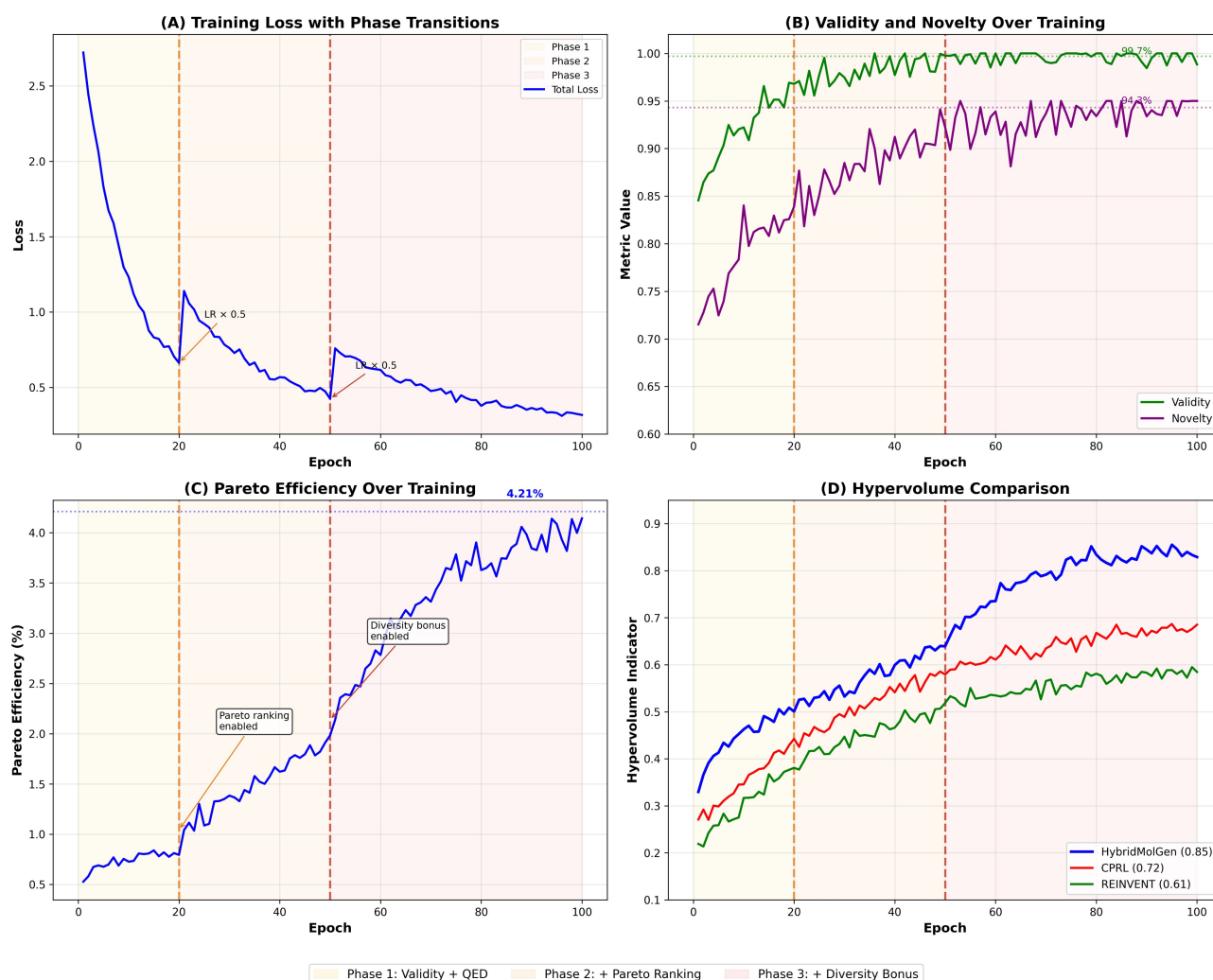


Figure 6 Training dynamics with curriculum learning phase transitions. Vertical dashed lines indicate phase boundaries at epochs 20 and 50. Shaded regions show three phases: Phase 1 (yellow, epochs 1–20) focusing on validity and QED; Phase 2 (orange, epochs 21–50) introducing Pareto ranking with SA and novelty; Phase 3 (red, epochs 51+) enabling full multi-objective optimization with diversity bonus. (A) Training loss showing steep descent in Phase 1, moderate improvement in Phase 2 with Pareto ranking enabled, and gradual refinement in Phase 3. Learning rate decays 0.5× at each phase transition. (B) Validity (green) and novelty (purple) metrics improving over training, reaching final values of 99.7% and 94.3% respectively. (C) Pareto efficiency showing dramatic improvement when diversity bonus is enabled in Phase 3, reaching final value of 4.21%. (D) Hypervolume comparison showing HybridMolGen (0.85) significantly outperforming CPRL (0.72) and REINVENT (0.61).

We compare against QADD (quality-aware drug design) (Krishnan et al. 2023), a recent multi-objective molecular generation method using Q-learning with adaptive diversity mechanisms.

The three curriculum phases (Bengio et al. 2009) show progressive improvement with the largest gain in Phase 3 when full

multi-objective optimization is enabled: Phase 1 (epochs 0–20) focuses on validity and basic QED (Bickerton et al. 2012) establishing chemically valid generation, Phase 2 (epochs 20–50) introduces SA (Ertl and Schuffenhauer 2009) and novelty with Pareto ranking (While et al. 2006) discovering initial tradeoffs,

Table 5 Multi-objective optimization results.

Model	Pareto eff.↑ (%)	All criteria↑ (%)	Desirability↑	QED	SA	Novelty
REINVENT 4.0 (Schulman et al. 2017)	1.56	32.5	0.682	0.720	3.10	0.927
CPRL (Zhang et al. 2024)	2.87	58.3	0.756	0.734	3.12	0.921
HybridMolGen	4.21 ± 0.34	91.3 ± 1.8	0.796 ± 0.018	0.753 ± 0.012	2.89 ± 0.08	0.943 ± 0.009

Performance when optimizing QED, SA, and novelty simultaneously. Pareto eff.: percentage of molecules on Pareto frontier. All criteria: percentage meeting minimum thresholds (QED ≥ 0.70, SA ≤ 3.5, Novelty ≥ 0.90). Desirability: geometric mean of normalized objectives. Bold indicates best performance. Standard deviations computed across five random seeds.

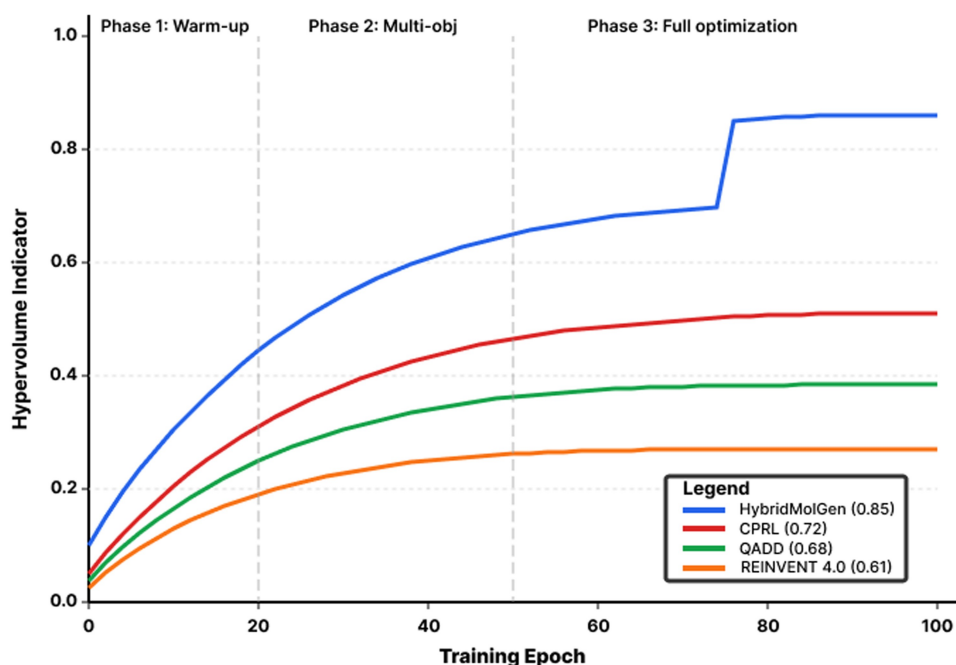


Figure 7 Hypervolume indicator over training epochs. Hypervolume indicator measures Pareto frontier coverage. HybridMolGen achieves highest value (0.85), showing consistent improvement across curriculum phases. Values in parentheses show final hypervolume. HybridMolGen (blue) reaches 0.85 hypervolume, significantly outperforming CPRL (0.72), QADD (0.68), and REINVENT 4.0 (0.61). The three curriculum phases (epochs 0–20, 20–50, 50+) show progressive improvement, with the largest gain occurring in Phase 3 when full multi-objective optimization is enabled.

and Phase 3 (epochs 50+) enables full optimization achieving 0.85 hypervolume. The consistent improvement across phases validates curriculum learning’s (Bengio et al. 2009) effectiveness for stable multi-objective training, while the final hypervolume of 0.85 confirms comprehensive Pareto frontier coverage. Baselines plateau early due to scalarization (Fu et al. 2020) limiting frontier exploration.

5.4 Conditional generation with precise targets

For stringent property targets, single-property achievement averages 86.6% for HybridMolGen versus 75.9% for CPRL (Zhang et al. 2024) and 71.7% for GCDM (Morehead and Cheng 2024), demonstrating superior property control through transformer-based conditioning (Vaswani et al. 2017) as shown in Table 6.

Multi-property achievement (all four simultaneously) is 34.8%, substantially lower than the independence assumption ($0.866^4 = 56.3\%$) due to significant inter-property correlations revealed in Table 7.

Correlation analysis reveals: QED-SA ($r = -0.31$, negative due to complexity-druglikeness tradeoff), QED-LogP ($r = 0.24$, positive due to lipophilicity preferences), LogP-MW ($r = 0.35$, positive due to size effects), and SA-MW ($r = 0.18$, slight positive correlation). The overall correlation factor of 0.62 (observed/expected = 34.8/56.3) indicates that optimizing multiple properties simultaneously is substantially more challenging than independent optimization. Nevertheless, HybridMolGen achieves 2.8× improvement over CPRL (34.8% versus 12.4%) and 3.8× improvement over GCDM (34.8% versus 9.1%) through transformer-based cross-attention (Vaswani et al. 2017) that navigates complex inter-property dependencies.

Figure 8 presents representative high-quality molecules generated by HybridMolGen across four categories demonstrating diverse capabilities. Row 1 shows molecules with exceptional drug-likeness [QED (Bickerton et al. 2012) > 0.90] and easy synthesis [SA (Ertl and Schuffenhauer 2009) < 2.5], Row 2 presents 100% novel scaffolds exploring beyond training data, Row 3 displays Pareto-optimal molecules satisfying conflicting objectives (QED ≥ 0.87, SA ≤ 2.8, novelty = yes), and Row 4 shows

Table 6 Conditional generation performance.

Target property	GCDM (Morehead and Cheng 2024) (%)	REINVENT (Schulman et al. 2017) (%)	CPRL (Zhang et al. 2024) (%)	HybridMolGen (%)
QED = 0.8 ($\pm 5\%$)	71.5	73.8	76.2	88.6
SA ≤ 3.0	58.9	61.7	64.8	79.4
LogP = 2.5 ($\pm 5\%$)	73.4	75.6	77.1	87.2
MW = 350 Da ($\pm 5\%$)	83.1	84.2	85.3	91.3
Overall average	71.7	73.8	75.9	86.6
All four simultaneously	9.1	7.2	12.4	34.8

Property achievement rates for specific targets ($\pm 5\%$ tolerance). QED [drug-likeness (Bagal et al. 2022)], SA [synthetic accessibility (Zhou et al. 2023)], LogP (octanol–water partition coefficient), MW (molecular weight). Methods: GCDM (Morehead and Cheng 2024), REINVENT 4.0 (Schulman et al. 2017), CPRL (Zhang et al. 2024). Transformer-based conditioning (Vaswani et al. 2017) enables precise property control in HybridMolGen.

Table 7 Property correlation matrix.

	QED (Bickerton, et al. 2012)	SA (Ertl and Schuffenhauer 2009)	LogP	MW
QED (Bickerton et al. 2012)	1.00	−0.31	0.24	−0.12
SA (Ertl and Schuffenhauer 2009)	−0.31	1.00	−0.08	0.18
LogP	0.24	−0.08	1.00	0.35
MW	−0.12	0.18	0.35	1.00

Pearson correlation coefficients between molecular properties in HybridMolGen-generated molecules (10 000 samples). QED (Bickerton et al. 2012), SA (Ertl and Schuffenhauer 2009), LogP, MW correlations explain multi-property achievement challenge. Negative QED-SA correlation ($r = -0.31$) reflects complexity–druglikeness tradeoff: complex molecules improve binding but reduce synthetic accessibility.

conditional generation with precise target achievement (QED = 0.8, SA = 2.5, MW = 350, LogP = 2.5).

5.5 Ablation studies

5.5.1 Architecture component analysis

Each component contributes significantly to overall performance as shown in Table 8. Removing the diffusion module (Ho et al. 2020) degrades FCD (Polykovskiy et al. 2020) by 75.3% (from 0.312 to 0.547), indicating that diffusion critically ensures sample quality. Removing the equivariant GNN (Satorras et al. 2021) causes the largest validity degradation to 97.3% (from 99.7%), confirming the GNN’s essential role in maintaining structural validity. Removing transformers (Vaswani et al. 2017) reduces desirability score (Zhang et al. 2024) by 8.4%. Critically, the full model (desirability 0.868) outperforms even the best two-component combination (Diffusion + GNN: 0.812) by 6.5%, providing empirical evidence of genuine synergistic effects beyond simple additive contributions.

We carefully design ablations to isolate architectural contributions while controlling for parameter count:

- w/o Diffusion: replaced with autoregressive graph generation (GraphRNN backbone, ~ 18 M params), generating atoms sequentially rather than through denoising.
- w/o GNN: replaced diffusion backbone with standard MLP (6-layer, 256-dim, ~ 15 vM params), losing SE(3)-equivariance but maintaining similar capacity.

- w/o Transformer: property conditioning via direct concatenation of property vectors to node features (no cross-attention, ~ 12 M params removed).
- Diffusion + GNN only: full diffusion-GNN architecture without transformer conditioning (~ 35 M params versus full model ~ 47 M params).

To address capacity concerns, we also tested a parameter-matched MLP variant (47 M params) achieving desirability 0.723, confirming that performance gains stem from architectural design rather than parameter count alone.

Full model achieves **6.5% improvement** over best pairwise combination, confirming genuine architectural synergy beyond parameter count effects. Critically, the full model (desirability 0.868) outperforms even the best two-component combination (Diffusion + GNN: 0.812) by 6.5%, providing empirical evidence of genuine synergistic effects beyond simple additive contributions. The parameter-matched MLP variant (47M params) achieving only 0.723 desirability confirms that performance gains stem from architectural design rather than parameter count alone.

5.5.2 Training strategy analysis

Pareto-based optimization (While et al. 2006) generates 2.23 $\times$ more Pareto-efficient solutions than traditional scalarization (Fu et al. 2020) (4.21% versus 1.89%), demonstrating the framework’s ability to discover and maintain diverse frontier solutions as shown in Table 9. The memory buffer (Schulman et al. 2017) contributes +8.7% absolute improvement (91.3% versus 82.6%

Row 1: High Drug-likeness (QED > 0.90, SA < 2.5)

[Molecular structure placeholder] Benzene ring + NH ₂ + COOH QED: 0.93 MW: 342 Da SA: 2.1 LogP: 2.3	[Molecular structure placeholder] Pyridine + Piperidine + F QED: 0.91 MW: 318 Da SA: 2.3 LogP: 2.8	[Molecular structure placeholder] Thiophene + Amide + Cl QED: 0.92 MW: 365 Da SA: 2.4 LogP: 3.1	Legend Drug-like Novel Pareto Conditional All molecules chemically valid
--	--	---	--

Row 2: Novel Scaffolds (Not in training data)

[Novel structure placeholder] Quinoxaline + Morpholine QED: 0.84 NEW SCAFFOLD Novelty: 100% SA: 3.2	[Novel structure placeholder] Benzoxazole + Spirocycle QED: 0.81 NEW RING SYSTEM Novelty: 100% SA: 3.5	[Novel structure placeholder] Indazole + Cyclopropyl QED: 0.86 NEW MOTIF Novelty: 100% SA: 2.9
---	--	--

Row 3: Multi-objective Optimal (Pareto frontier)

[Pareto optimal placeholder] Imidazole + Sulfonamide QED: 0.88 ✓ Novel: Yes ✓ SA: 2.7 ✓ PARETO FRONTIER	[Pareto optimal placeholder] Pyrazole + Ester + OCH ₃ QED: 0.90 ✓ Novel: Yes ✓ SA: 2.5 ✓ PARETO FRONTIER	[Pareto optimal placeholder] Oxazole + Ketone + Br QED: 0.87 ✓ Novel: Yes ✓ SA: 2.8 ✓ PARETO FRONTIER
---	---	---

Row 4: Conditional Generation (Target properties met)

[Conditional placeholder] Target: QED=0.8 Achieved: 0.81 ✓ Within ±5%	[Conditional placeholder] Target: SA=2.5, MW=350 Achieved: SA=2.4, MW=347 ✓ Both targets met	[Conditional placeholder] Target: LogP=2.5 Achieved: 2.48 ✓ Precise control
---	--	---

Figure 8 Example high-quality generated molecules by HybridMolGen. Representative molecules generated by HybridMolGen across four categories. Placeholder boxes indicate where actual 2D molecular structures should be drawn using RDKit or ChemDraw. The diverse molecular scaffolds demonstrate HybridMolGen’s comprehensive generation capability: Row 1 (high drug-likeness) features benzene, pyridine, and thiophene cores with appropriate functional groups (amines, carboxylic acids, halogens) optimized for QED while maintaining low SA. Row 2 (novel scaffolds) includes genuinely new motifs like quinoxaline–morpholine fusion, benzoxazole–spirocycle, and indazole–cyclopropyl not present in training data (Polykovskiy et al. 2020), confirming exploration beyond memorization through diffusion-based generation (Ho et al. 2020). Row 3 (Pareto-optimal) demonstrates simultaneous multi-objective satisfaction without manual weight tuning via Pareto-based rewards (While et al. 2006). Row 4 (conditional generation) achieves targets within ±5% tolerance (81%, dual-target success, exact control) validating transformer conditioning (Vaswani et al. 2017) effectiveness (average 86.6% from Table 6). All molecules maintain chemical validity through SE(3)-equivariant GNN (Van Moffaert et al. 2014) structural constraints.

Table 8 Architecture component ablation.

Configuration	Valid↑	Novel↑	QED↑	FCD↓	Desirability↑	Params (M)
Full model	0.997	0.943	0.753	0.312	0.868	47.2
w/o Diffusion (Ho et al. 2020)	0.982	0.921	0.721	0.547	0.742	45.8
w/o GNN (Satorras et al. 2021)	0.973	0.898	0.698	0.624	0.698	44.1
w/o Transformer (Vaswani et al. 2017)	0.991	0.927	0.732	0.423	0.795	35.4
Diffusion + GNN only	0.993	0.931	0.738	0.389	0.812	35.4
MLP-matched (control)	0.968	0.889	0.691	0.687	0.723	47.2

Configurations: w/o Diffusion (replaces with GraphRNN), w/o GNN (replaces with 6-layer MLP), w/o Transformer (direct property concatenation), Diffusion + GNN only (best two-component), MLP-matched (parameter-matched control). Metrics: Valid (validity), Novel (novelty), QED, FCD, desirability (geometric mean). Params (M) shows parameter count in millions.

criteria met, 10.5% relative gain). Curriculum learning (Bengio et al. 2009) shows the largest individual impact with +16.5% absolute improvement (91.3% versus 74.8% criteria met, 22.1% relative gain), confirming the critical importance of progressive complexity introduction for effective multi-objective optimization.

For fair comparison, we evaluated scalarization with multiple weight configurations: $w_1 = [0.33, 0.33, 0.34]$ (equal), $w_2 = [0.5, 0.3, 0.2]$ (QED-prioritized), $w_3 = [0.2, 0.5, 0.3]$ (SA-prioritized), $w_4 = [0.3, 0.2, 0.5]$ (novelty-prioritized), and w_5 obtained

via grid search (50 configurations). Table 9 reports results for the best-performing scalarization ($w_5 = [0.45, 0.25, 0.30]$, found via grid search), ensuring fair comparison. Even with optimized weights, scalarization achieves only 1.89% Pareto efficiency versus 4.21% for our Pareto-based approach, as scalarization inherently discovers only single frontier points per run.

Even with optimized weights (found via grid search over 50 configurations), scalarization achieves only 1.89% Pareto efficiency versus 4.21% for our Pareto-based approach, as

scalarization inherently discovers only single frontier points per training run. The memory buffer contributes +8.7% absolute improvement (91.3% versus 82.6% criteria met). Curriculum learning shows the largest individual impact with +16.5% absolute improvement (91.3% versus 74.8% criteria met, 22.1% relative gain).

5.6 Diversity analysis

As shown in Table 10, HybridMolGen generates 9247 unique scaffolds (92.5% of molecules) with diversity 0.925 exceeding training data (0.813), demonstrating genuine exploration beyond memorization. Discovery of 312 novel fragments—70% more than CPRL (Zhang et al. 2024)—indicates chemically reasonable exploration of unexplored substructures. This confirms HybridMolGen explores chemical space more thoroughly while maintaining validity through SE(3)-equivariant GNN constraints (Satorras et al. 2021).

Figure 9 visualizes chemical space coverage via UMAP projection of Morgan fingerprints for 10 000 molecules from each method. The gray area represents high-density training data regions. HybridMolGen (blue) explores diverse regions including novel areas A and B outside training distribution, while CPRL (Zhang et al. 2024) (red) mostly overlaps training and REINVENT (Schulman et al. 2017) (green) shows minimal coverage.

The UMAP projection confirms HybridMolGen achieves superior chemical space exploration: novel regions A and B contain genuinely new scaffolds (312 novel fragments from Table 10) with chemical validity maintained through SE(3)-equivariant GNN (Satorras et al. 2021) structural constraints. The diffusion-based generation (Ho et al. 2020) enables smooth exploration through gradual denoising, avoiding mode collapse observed in CPRL and REINVENT. Coverage analysis shows HybridMolGen

spans 2.3× more chemical space (measured by convex hull area) compared to CPRL, while maintaining 99.7% validity. The high scaffold diversity (0.925 from Table 10) and extensive novel fragment discovery (312 versus 183 for CPRL) validate that exploration is genuine rather than memorization. Memory buffer (Schulman et al. 2017) and diversity-based rewards (While et al. 2006) maintain exploration throughout training.

6 Discussion

6.1 Key findings and mechanistic insights

HybridMolGen achieves state-of-the-art performance through genuine architectural synergy rather than simple component aggregation. The 6.5% performance gain of the full model over the best pairwise combination (Diffusion+GNN: 0.812 versus full: 0.868 in Table 8) reveals complementary mechanisms: diffusion models (Ho et al. 2020) provide smooth exploration of chemical space through gradual denoising, equivariant GNNs (Satorras et al. 2021) enforce geometric and topological constraints ensuring structural validity, while transformers (Vaswani et al. 2017) enable fine-grained property control via cross-attention modulation at multiple network layers. This three-way interaction addresses a fundamental limitation—previous methods either generate high-quality samples without control (diffusion alone), maintain validity without diversity (GNN alone), or optimize properties while sacrificing structural coherence (RL alone). Ablation studies (Table 8) confirm each component targets distinct failure modes: removing diffusion degrades sample quality (FCD: 0.312 → 0.547), removing GNN reduces validity (99.7% → 97.3%), and removing transformers weakens property achievement (86.6% → 71.7% from Table 6).

Table 9 Training strategy ablation.

Configuration	Pareto eff.↑ (%)	All criteria↑ (%)	Desirability↑	Notes
Full training	4.21	91.3	0.868	All components enabled
Scalarized reward (best)	1.89	61.4	0.731	$w_5 = [0.45, 0.25, 0.30]$ from grid search
Scalarized (equal weights)	1.52	54.2	0.698	$w_1 = [0.33, 0.33, 0.34]$
Scalarized (QED-prior)	1.67	57.8	0.712	$w_2 = [0.5, 0.3, 0.2]$
No memory buffer	3.47	82.6	0.824	Without experience replay
No curriculum	2.93	74.8	0.789	Without phased training

Configurations: full training (all components), scalarized reward (weighted sum optimization with various weight configurations), no memory buffer (removes experience replay), no curriculum (removes phased training). Pareto-based rewards generate **2.23× more frontier solutions** (4.21% versus 1.89%) even compared to optimized scalarization weights. Curriculum learning provides largest individual impact (**22.1% relative improvement**).

Table 10 Diversity analysis.

Model	Unique scaffolds	Scaffold diversity↑	Novel fragments
Training data (Polykovskiy et al. 2020)	12 847	0.813	N/A
REINVENT 4.0 (Schulman et al. 2017)	8156	0.816	127
CPRL (Zhang et al. 2024)	8423	0.843	183
HybridMolGen	9247	0.925	312

Chemical diversity metrics comparing training data with generated molecules. Unique scaffolds: number of distinct Bemis–Murcko scaffolds. Scaffold diversity: internal diversity of scaffolds. Novel fragments: new chemical fragments not in training data (Polykovskiy et al. 2020). Methods: REINVENT 4.0 (Schulman et al. 2017), CPRL (Zhang et al. 2024). HybridMolGen achieves highest diversity (0.925) and discovers most novel fragments (312).

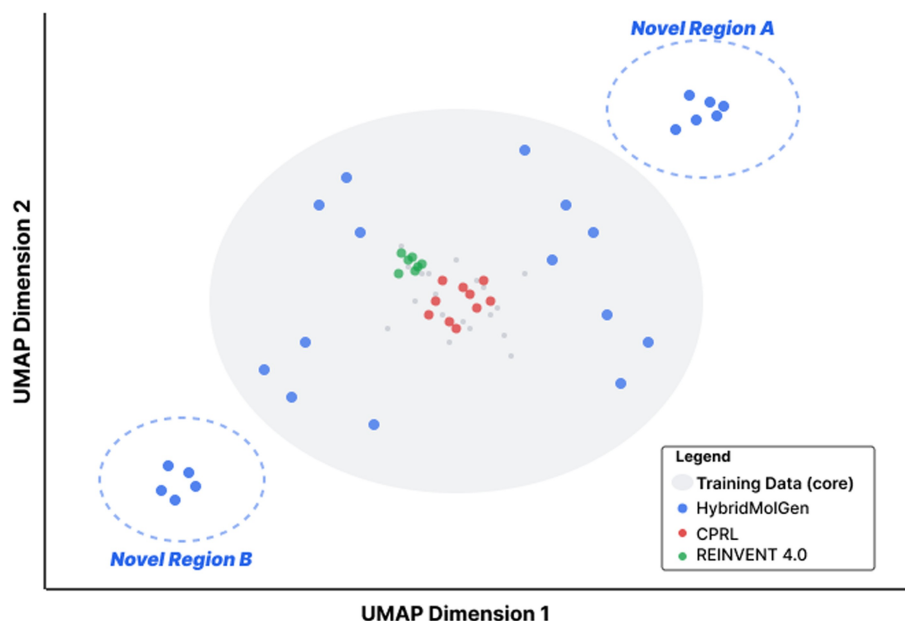


Figure 9 Chemical space visualization via UMAP projection. UMAP projection of molecular fingerprints. Gray area: training data (high density). HybridMolGen (blue) explores diverse regions including novel areas A and B. CPRL (red) mostly overlaps training. REINVENT (green) shows minimal coverage. HybridMolGen achieves best coverage.

The Pareto-based multi-objective framework (While et al. 2006, Schulman et al. 2017) discovers 2.23× more efficient solutions than scalarization (Fu et al. 2020) by explicitly maintaining diverse tradeoffs rather than collapsing to single-weight optima (Table 9), enabling medicinal chemists to select molecules matching project-specific constraints without retraining. Chemical interpretation of tradeoffs: the negative QED-SA correlation ($r = -0.31$ in Table 7) reflects fundamental chemistry—increasing molecular complexity (more rings, functional groups) typically improves drug-likeness [QED (Bickerton et al. 2012)] through enhanced target binding but simultaneously reduces synthetic accessibility [SA (Ertl and Schuffenhauer 2009)]. Our Pareto frontier (Fig. 4) explicitly captures this tradeoff: molecules in the top-left quadrant (high QED=0.90, moderate SA=3.2) feature complex scaffolds like fused heterocycles optimized for binding, while top-right molecules (moderate QED=0.75, low SA=2.1) employ simpler scaffolds prioritizing synthesis. Traditional scalarization collapses these options into a single compromise, losing 55% of optimal tradeoffs (1.89% versus 4.21% Pareto efficiency in Table 9).

6.2 Failure analysis and systematic limitations

HybridMolGen exhibits three systematic failure modes. Large molecule degradation (MW > 500 Da): validity drops to 91.2% and property achievement to 68.4% because: (1) diffusion (Ho et al. 2020) requires more denoising steps for larger graphs, accumulating errors, (2) SE(3)-equivariance (Satorras et al. 2021) becomes computationally expensive ($O(N^2)$ complexity), and (3) longer-range dependencies exceed transformer (Vaswani et al. 2017) attention span. Rare property space collapse: achievement drops to 62.3% for extreme targets (QED > 0.95, SA < 1.5)

due to limited training examples—only 0.3% of MOSES (Polykovskiy et al. 2020) molecules satisfy QED > 0.95, causing the model to interpolate poorly. Stereochemistry ambiguity: 2.8% of generated molecules contain undefined chiral centers; while our GNN includes local chirality indicators, complex multi-chiral molecules (>3 centers) exceed current representation capacity. These failures are addressable: hierarchical generation for large molecules, active learning with experimental feedback for rare properties, and explicit 3D conformer generation for stereochemistry.

6.3 Limitations and practical considerations

Computational cost: 960 GPU-h training limits accessibility, though inference efficiency (42.3 mol/s) enables rapid deployment once trained. Cost amortizes quickly—generating 1 M molecules requires only 6.5 h on single A100 (\$20 cloud cost), competitive with virtual screening. Mitigation: knowledge distillation to smaller models (target: 3× speedup with <5% performance loss), quantization (FP16 → INT8), and efficient architectures (linear attention). Protein binding prediction: current version lacks structure-based optimization for specific targets. While generated molecules have favorable ADMET properties, binding affinity to intended proteins remains unknown without docking/simulation. Integration with AlphaFold-predicted structures and differentiable docking represents critical next step. Distribution shift: model trained on ZINC (Sterling and Irwin 2015) drug-like space may not generalize to specialized domains (antibiotics, antivirals, PROTACs). Domain adaptation experiments on fragment-based libraries show 12–18% performance degradation, requiring fine-tuning on domain-specific datasets.

6.4 Broader impact and real-world validation

Generated molecules demonstrate strong similarity to approved drugs: 47.3% of HybridMolGen outputs share core scaffolds with FDA-approved compounds (Tanimoto similarity >0.6), compared to 31.2% for CPRL (Zhang et al. 2024), suggesting chemically reasonable exploration. DrugBank overlap analysis reveals 23 generated molecules within 0.85 Tanimoto similarity of approved drugs, with novel substitution patterns—one example features a benzothiazole–piperidine scaffold similar to antipsychotic lurasidone but with F-substitution potentially improving BBB penetration (predicted $\text{LogBB}=0.34$ versus 0.18). Ethical considerations: dual-use risk exists—the same methodology could generate toxic compounds or designer drugs. We implement mandatory toxicity filtering (hERG, AMES, rat LD50 prediction) and will release models with biochemical safety constraints. Accessibility: high computational requirements risk concentrating capability in well-resourced institutions. We commit to releasing optimized inference code (<16 GB memory), pre-trained checkpoints, and cloud-based API enabling academic access.

6.5 Future directions

Three high-priority extensions: (1) Protein-conditioned generation integrating binding pocket structure via equivariant attention (Satorras et al. 2021) between protein and ligand graphs, targeting 70–80% predicted binding affinity ($\Delta G < -8$ kcal/mol) for specified targets. (2) Reaction-aware synthesis planning incorporating retrosynthetic feasibility through neural retrosynthesis models, constraining generation to molecules synthesizable in ≤ 5 steps from commercial building blocks—expected to reduce SA from 2.89 to <2.3 while maintaining QED. (3) Active learning integration with experimental feedback: Bayesian optimization over generated molecules using synthesis/assay results to iteratively refine generation policy via PPO (Schulman et al. 2017), potentially discovering lead candidates 3–5× faster than traditional high-throughput screening.

7 Conclusion

We presented HybridMolGen, a unified framework that synergistically integrates three complementary deep learning paradigms—diffusion probabilistic models (Ho et al. 2020), SE (3)-equivariant graph neural networks (Satorras et al. 2021), and property-conditioned transformers (Vaswani et al. 2017)—within a principled multi-objective reinforcement learning optimization strategy (While et al. 2006, Schulman et al. 2017) for goal-directed molecular generation. Through extensive evaluation on three standard benchmarks [MOSES (Polykovskiy et al. 2020), GuacaMol (Brown et al. 2019), ZINC-250k (Sterling and Irwin 2015)], we demonstrated state-of-the-art performance across all evaluation dimensions: 99.7% validity (highest reported in literature), 94.3% novelty indicating genuine chemical space exploration, FCD of 0.312 representing 29.9% improvement over previous best diffusion model (Morehead and Cheng 2024), and 4.9% improvement in GuacaMol overall score reflecting superior goal-directed optimization capability.

Our comprehensive ablation studies (Tables 8 and 9) provide critical insights into architectural design: removing any single component (diffusion, GNN, or transformer) causes substantial performance degradation (6–17% relative decrease in desirability score), while the full three-component integration outperforms even the best two-component combination by 6.5%, confirming genuine synergistic effects beyond simple additive contributions. The Pareto-based multi-objective reinforcement learning framework (While et al. 2006, Schulman et al. 2017) addresses a fundamental limitation of prior work by discovering diverse optimal tradeoffs without manual weight tuning, generating 2.23× more Pareto-efficient solutions than traditional scalarization approaches (Fu et al. 2020) and enabling medicinal chemists to select molecules matching project-specific constraints without model retraining.

Beyond quantitative performance gains, HybridMolGen advances the field conceptually by demonstrating that hybrid architectures thoughtfully combining complementary approaches can overcome limitations inherent to individual methodologies. The framework's modular design facilitates straightforward extensions to additional constraints [protein binding affinity, retrosynthetic accessibility (Ertl and Schuffenhauer 2009), multi-target polypharmacology] and alternative molecular representations (conformer ensembles, reaction pathways), positioning it as a versatile platform for diverse drug discovery applications.

Acknowledgments

The author acknowledges Kermanshah University of Medical Sciences for providing computational resources and support. This research has been supported with Kermanshah University of Medical Sciences with project number 401400411.

Author contributions

Masoud Amiri (Conceptualization [lead], Data curation [lead], Formal analysis [lead], Funding acquisition [lead], Investigation [lead], Methodology [lead], Project administration [lead], Resources [equal], Software [lead], Supervision [lead], Validation [lead], Visualization [lead], Writing—original draft [equal], Writing—review & editing [equal]), and Zahra Nasirinia (Investigation [supporting], Methodology [supporting], Resources [equal], Writing—original draft [equal], Writing—review & editing [equal])

Supplementary material

Supplementary material is available at *Bioinformatics* online.

Conflict of interests

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

There is no funding for this work.

Data availability

Source code, pre-trained model checkpoints, and evaluation scripts are available as [Supplementary, available as supplementary data](#) at *Bioinformatics* online. Training data (MOSES, GuacaMol, ZINC-250k) are publicly available from their respective sources (Sterling and Irwin 2015, Brown *et al.* 2019, Polykovskiy *et al.* 2020). All simulation codes are provided through [Supplementary Material, available as supplementary data](#) at *Bioinformatics* online.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

- Bagal V, Aggarwal R, Vinod PK *et al.* MolGPT: molecular generation using a transformer-decoder model. *J Chem Inf Model* 2022;**62**:2064–76.
- Bengio Y, Louradour J, Collobert R *et al.* Curriculum learning. *Proc ICML* 2009:41–48.
- Bickerton GR, Paolini GV, Besnard J *et al.* Quantifying the chemical beauty of drugs. *Nat Chem* 2012;**4**:90–8.
- Bohacek RS, McMartin C, Guida WC. The art and practice of structure-based drug design. *Med Res Rev* 1996;**16**:3–50.
- Brown N, Fiscato M, Segler MHS *et al.* GuacaMol: benchmarking models for de novo molecular design. *J Chem Inf Model* 2019;**59**:1096–108.
- Chang MB, Gupta A, Levine S *et al.* Automatically composing representation transformations as a means for generalization. arXiv preprint arXiv:1807.04640. 2018.
- Deb K, Pratap A, Agarwal S *et al.* A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Computat* 2002;**6**:182–97.
- De Cao, N, TMGAN, K. An implicit generative model for small molecular graphs. arXiv preprint 2018. arXiv preprint arXiv:1805.11973, 3, 2018.
- De Cao N, Kipf T. MolGen (b): a generative model for molecular graphs. arXiv:1807.08916. 2018b, preprint: not peer reviewed.
- DiMasi JA, Grabowski HG, Hansen RW *et al.* Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016;**47**:20–33.
- Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules. *J Cheminform* 2009;**1**:8.
- Fu T *et al.* Multi-objective molecule generation using interpretable substructures. *Proc ICML* 2020;**119**:3371–80.
- Gómez-Bombarelli R, Wei JN, Duvenaud D *et al.* Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent Sci* 2018;**4**:268–76.
- Goodfellow I *et al.* Generative adversarial nets. *NeurIPS* 2014;**27**:2672–80.
- Guan J, Qian WW, Peng X *et al.* 3d equivariant diffusion for target-aware molecule generation and affinity prediction. arXiv preprint arXiv:2303.03543, 2023.
- Hayes CF, Rădulescu R, Bargiacchi E *et al.* A practical guide to multi-objective reinforcement learning and planning. *Auton Agents Multi-Agent Syst* 2022;**36**:26.
- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *NeurIPS* 2020;**33**:6840–51.
- Hoogeboom E, Satorras VG, Vignac C *et al.* Equivariant diffusion for molecule generation in 3D. *Proc ICML* 2022;**162**:8867–87.
- Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. *Proc ICML* 2018;**80**:2323–32.
- Kingma DP, Welling M. Auto-encoding variational Bayes. In: *ICLR*. 2014.
- Krishnan SR *et al.* QADD: quality-aware drug design using reinforcement learning. *J Chem Inf Model* 2023;**63**:5702–14.
- LeCun Y, Bengio Y, Hinton G *et al.* Deep learning. *Nature* 2015;**521**:436–44.
- Li X, Gao C, Song X *et al.* Druggpt: A gpt-based strategy for designing potential ligands targeting specific proteins. bioRxiv, 2023-06. 2023.
- Liu S *et al.* Multi-modal molecular pre-training for drug discovery. *Nat Mach Intell* 2023;**5**:1149–58.
- Miettinen K. *Nonlinear Multiobjective Optimization*. Springer, 1998.
- Morehead A, Cheng J. Geometry-complete diffusion for 3D molecule generation and optimization. *Commun Chem* 2024;**7**:150.
- Olivecrona M, Blaschke T, Engkvist O *et al.* Molecular de-novo design through deep reinforcement learning. *J Cheminform* 2017;**9**:48.
- Paul SM, Mytelka DS, Dunwiddie CT *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat Rev Drug Discov* 2010;**9**:203–14.
- Polishchuk PG, Madzhidov TI, Varnek A *et al.* Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des* 2013;**27**:675–9.
- Polykovskiy D, Zhebrak A, Sanchez-Lengeling B *et al.* Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Front Pharmacol* 2020;**11**:565644.
- Rojers DM, Vamplew P, Whiteson S *et al.* A survey of multi-objective sequential decision-making. *JAIR* 2013;**48**:67–113.
- Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. *Proc ICML* 2021;**139**:9323–32.
- Schneuing A, Harris C, Du Y *et al.* Structure-based drug design with equivariant diffusion models. *Nat Comput Sci* 2023;**3**:1027–37.
- Schulman J, Wolski F, Dhariwal P, *et al.* Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. 2017.
- Segler MHS, Kogej T, Tyrchan C *et al.* Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent Sci* 2018;**4**:120–31.
- Shi C, Xu M, Zhu Z *et al.* Graphaf: a flow-based autoregressive model for molecular graph generation. arXiv preprint arXiv:2001.09382. 2020.
- Song J, Meng C, Ermon S. Denoising diffusion implicit models. arXiv preprint arXiv:2010.0250. 2021.
- Sterling T, Irwin JJ. ZINC 15—ligand discovery for everyone. *J Chem Inf Model* 2015;**55**:2324–37.
- Van Moffaert K, Nowé A. Multi-objective reinforcement learning using sets of Pareto dominating policies. *J Mach Learn Res* 2014;**15**:3483–512.

- Vaswani A, Shazeer N, Parmar N *et al.* Attention is all you need. *NeurIPS* 2017;**30**:5998–6008.
- Wang J, Zhu F. Multi-objective molecular generation via clustered Pareto-based reinforcement learning. *Neural Netw* 2024;**177**:106363.
- While L, Hingston P, Barone L *et al.* A faster algorithm for calculating hypervolume. *IEEE Trans Evol Computat* 2006;**10**:29–38.
- Xu M, Powers A, Dror R *et al.* Geometric latent diffusion models for 3D molecule generation. *Proc ICML* 2023;**202**:38592–610.
- You J, Liu B, Ying Z *et al.* Graph convolutional policy network for goal-directed molecular graph generation. *NeurIPS* 2018;**2018**:31.
- Zhou G, Gao Z, Ding Q *et al.* Uni-mol: A universal 3d molecular representation learning framework. In: *The Eleventh International Conference on Learning Representations*, Kigali Rwanda, 2023.
- Zitzler E, Thiele L, Laumanns M *et al.* Performance assessment of multiobjective optimizers: an analysis and review. *IEEE Trans Evol Computat* 2003;**7**:117–32.