CAGenMol: Condition-Aware Diffusion Language Model for Goal-Directed Molecular Generation

Yanting LI[1,*], Zhuoyang JIANG[1*], Enyan DAI[1], Lei WANG[2], Wen-Cai Ye[2], Li LIU[1],
[1] The Hong Kong University of Science and Technology (Guangzhou),
[2]Jinan University, Guangzhou
Correspondence: avrillliu@hkust-gz.edu.cn

Abstract

Goal-directed molecular generation requires satisfying heterogeneous constraints such as protein–ligand compatibility and multi-objective drug-like properties, yet existing methods often optimize these constraints in isolation, failing to reconcile conflicting objectives (e.g., affinity vs. safety), and struggle to navigate the non-differentiable chemical space without compromising structural validity. To address these challenges, we propose CAGenMol, a condition-aware discrete diffusion framework over molecular sequences that formulates molecular design as conditional denoising guided by heterogeneous structural and property signals. By coupling discrete diffusion with reinforcement learning, the model aligns the generation trajectory with non-differentiable objectives while preserving chemical validity and diversity. The non-autoregressive nature of diffusion language model further enables iterative refinement of molecular fragments at inference time. Experiments on structure-conditioned, property-conditioned, and dual-conditioned benchmarks demonstrate consistent improvements over state-of-the-art methods in binding affinity, drug-likeness, and success rate, highlighting the effectiveness of our framework. The code is available at https://github.com/Lee612-1/CAGenMol.

1 Introduction

The discovery of novel small-molecule therapeutics is a cornerstone of modern medicine (DiMasi et al., 2016; Hughes et al., 2011). However, a clinical candidate must simultaneously satisfying diverse application settings and their corresponding constraints (Polishchuk et al., 2013; Ferreira and Andricopulo, 2019). This multi-objective nature makes the search space vast and discontinuous, rendering traditional trial-and-error approaches inefficient. Consequently, generative models have emerged as a promising paradigm to accelerate goal-directed molecular design.

Despite their promise, existing approaches typically compromise between structural precision and optimization flexibility. Structure-Based Drug Design (SBDD) methods directly model 3D atomic interactions to ensure high binding affinity (Peng et al., 2022; Guan et al., 2023). However, they rely on computationally expensive 3D representing and struggle to optimize non-geometric pharmacological properties. Conversely, sequence-based optimization methods treat molecular generation as a black-box search (Loeffler et al., 2024; Zhou et al., 2019). While flexible in defining objectives, they often lack structural priors, leading to chemically invalid results (Renz et al., 2019). Therefore, a unified framework that reconciles structural perception with robust multi-objective optimization remains an open challenge.

From a modeling perspective, 1D sequence representations (e.g., SMILES) offer a computationally efficient alternative to 3D conformations. However, the dominant Autoregressive (AR) models (Bagal et al., 2021; Noutahi et al., 2024) generate tokens in a rigid left-to-right order. This mechanism fundamentally limits their ability to incorporate global structural contexts or perform local refinement, making them fragile when coupled with aggressive RL policy. To address these limitations, we identify Discrete Diffusion Language Models (DLMs) as a superior substrate for goal-directed generation. DLMs generate molecules via a non-autoregressive iterative denoising process (Austin et al., 2021; Sahoo et al., 2024). This paradigm offers two implicit advantages: (1) Global Visibility, which allows the model to attend to conditioning signals simultaneously across the entire sequence; and (2) Iterative Editability, which permits fine-grained structural corrections during generation.

[*]These authors contributed equally to this work.

Findings of the Association for Computational Linguistics: ACL 2026, pages 4720–4738
July 2-7, 2026

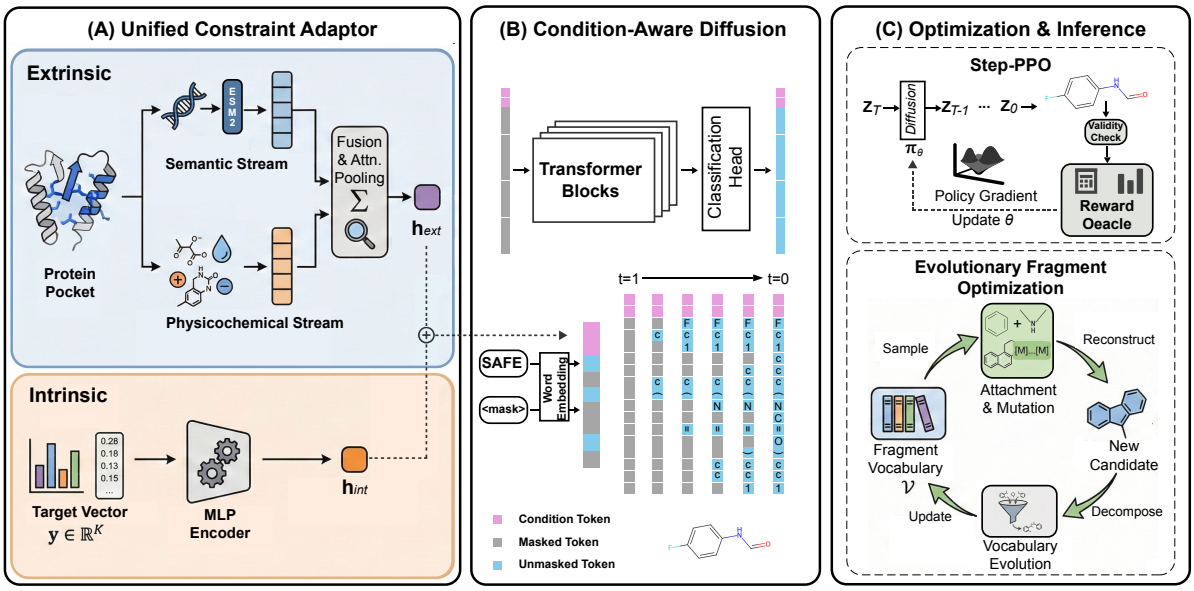

Figure 1: Overview of CAGenMol. UCA encodes either protein-pocket structure or target properties, which guides a condition-aware masked diffusion model on SAFE sequences. The model is further trained with Step-PPO, and perform inference with EFO refinement.

Despite these potentials, how to effectively ground DLMs on complex biological constraints and optimize them via reinforcement learning in discrete space remains underexplored.

In this work, we propose CAGenMol, a unified framework that synergizes condition-aware discrete diffusion with reinforcement learning to bridge the gap between biological constraints and chemical space. First, to handle heterogeneous inputs, we introduce a **Unified Constraint Adaptor (UCA)** that projects diverse signals into a shared latent space, serving as a persistent semantic anchor for the diffusion backbone. Building on diffusion process, we develop **Step-wise Proximal Policy Optimization (Step-PPO)**. Unlike token-level RL on AR model, Step-PPO leverages the iterative nature of diffusion to perform fine-grained credit assignment at each denoising step, ensuring precise alignment with complex objectives without collapsing the generative prior. Finally, to surpass the limitations of one-shot sampling, we propose **Evolutionary Fragment Optimization (EFO)**, an inference-time refinement strategy that exploits the editing flexibility of our model to perform gradient-free hill-climbing on generated candidates.

Our contributions are summarized as follows: (1) We formulate goal-directed molecular generation as a conditional discrete diffusion problem, providing a unified modeling perspective that naturally accommodates heterogeneous structural and property constraints. (2) We establish a diffusion-aware optimization framework that enables step-level policy learning within the denoising process, allowing effective alignment with complex objectives in discrete chemical space. (3) We propose a principled inference-time refinement mechanism that leverages the flexibility of non-autoregressive diffusion models to iteratively improve generated molecules while preserving diversity. (4) Comprehensive experiments across multiple conditional generation settings validate the effectiveness and robustness of the proposed framework.

2 Related Work

Diffusion Language Models Transformer architectures (Vaswani et al., 2017) have revolutionized sequence modeling. While autoregressive models such as GPT (Radford et al., 2018) achieve strong performance, their left-to-right decoding results in slow inference and limited editability. Discrete Diffusion Language Models (DLMs) mitigate these issues through non-autoregressive generation. By introducing diffusion processes over discrete token spaces (Austin et al., 2021; Lou et al., 2024; Sahoo et al., 2024), DLMs enable faster generation, greater diversity, and flexible token ordering. Recent work such as LLaDA-V (You et al., 2025) extends DLMs to multimodal settings; however, their application to scientific domains requiring precise conditioning on heterogeneous biological signals remains largely unexplored.

Structure-Based Drug Design Structure-Based Drug Design(SBDD) aims to generate ligands that specifically bind to a target protein pocket. Existing methods (Peng et al., 2022; Guan et al., 2023; Jain et al., 2023) typically model the joint 3D atomic distribution, with recent work (Qu et al., 2024; Zhang et al., 2025) further incorporating latent diffusion or geometric vector quantization. Despite their success, these approaches face two key limitations: they rely on computationally expensive explicit 3D representations that assume static protein structures, and they primarily optimize geometric or binding criteria while neglecting broader drug-like properties such as ADMET and synthetic accessibility.

Optimization-Based Molecular Generation Goal-directed molecular generation is commonly formulated as an optimization problem, with Reinforcement Learning (RL) (Loeffler et al., 2024; Zhou et al., 2019; Wang et al., 2024)and Genetic Algorithms (GA) (Yoshikawa et al., 2018; Spiegel and Durrant, 2020; Jensen, 2019) as the dominant paradigms which respectively optimize generative policies toward target objectives and evolve molecular populations via mutation and crossover. While both paradigms support flexible objective definitions and thus offer strong generality, they typically treat chemical space as a black box, relying on weak generative priors and sparse reward signals. As a result, RL-based methods are prone to reward hacking and mode collapse, whereas GA-based approaches often generate chemically invalid molecules.

3 Preliminary

3.1 Problem Definition

We formulate goal-directed molecular generation as a conditional sequence generation task, aiming to learn a distribution $p_\phi(\mathcal{S} \mid \mathbf{c_s}, \mathbf{c_p})$ over molecular sequences $\mathcal{S} = [s_1, \ldots, s_L]$, where $\mathbf{c_s}, \mathbf{c_p}$ are:

Extrinsic Structural Condition. $\mathbf{c}_s$ represents a 3D protein pocket $\mathcal{P} = (\mathbf{r}_i, \mathbf{v}_i)_{i=1}^N$, where $\mathbf{r}_i$ and $\mathbf{v}_i$ denote atomic coordinates and chemical features. The goal is to generate ligands that bind $\mathcal{P}$ with high affinity.

Intrinsic Property Condition. $\mathbf{c}_p$ specifies a K-dimensional target property vector $\mathbf{y} \in \mathbb{R}^K$, and the objective is to generate molecules whose properties match $\mathbf{y}$. In both cases, $p\phi$ must generate valid molecular structures that satisfy the given condition $\mathbf{c}$.

3.2 Diffusion Language Model

Masked diffusion models (Sahoo et al., 2024; Shi et al., 2024) are an effective class of diffusion language models. We follow MDLM (Sahoo et al., 2024) to define the masked diffusion process.

Let $\boldsymbol{x}$ be a length-L sequence, where each token $\boldsymbol{x}^l$ is a one-hot vector over K categories, satisfying $\boldsymbol{x}_i^l \in \{0,1\}^K$ and $\sum_{i=1}^K \boldsymbol{x}_i^l = 1$. The K-th category corresponds to the masking token $\mathbf{m}$, with $\mathbf{m}_K = 1$. We denote by $\mathrm{Cat}(\cdot; \boldsymbol{\pi})$ a categorical distribution with parameter $\boldsymbol{\pi} \in \Delta^K$.

The forward process gradually replaces clean tokens with the mask according to

$$q(\boldsymbol{z}_t^l \mid \boldsymbol{x}^l) = \mathrm{Cat}\big(\boldsymbol{z}_t^l;\, \alpha_t \boldsymbol{x}^l + (1-\alpha_t)\mathbf{m}\big)\,, \qquad (1)$$

where $\boldsymbol{z}_t^l$ denotes the l-th token at time $t \in [0,1]$, and α_t is a monotonically decreasing masking schedule with $\alpha_0 = 1$ and $\alpha_1 = 0$. At $t = 1$, all tokens are masked.

The reverse process recovers less-masked sequences from more-masked ones. For $s < t$, the reverse transition $p_\theta(\boldsymbol{z}_s^l \mid \boldsymbol{z}_t^l)$ is defined as

$$\begin{cases} \mathrm{Cat}(\boldsymbol{z}_s^l; \boldsymbol{z}_t^l), & \boldsymbol{z}_t^l \neq \mathbf{m}, \\ \mathrm{Cat}\left(\boldsymbol{z}_s^l;\, \dfrac{(1-\alpha_s)\mathbf{m} + (\alpha_s - \alpha_t)\,\boldsymbol{x}_\theta^l(\boldsymbol{z}_t, t)}{1-\alpha_t}\right), & \boldsymbol{z}_t^l = \mathbf{m}, \end{cases} \qquad (2)$$

where $\boldsymbol{x}_\theta(\boldsymbol{z}_t, t)$ is a denoising network predicting the clean-token distributions. This formulation preserves already unmasked tokens.

3.3 SAFE and Base Model

We adopt SAFE (Noutahi et al., 2024) to align with the non-autoregressive nature of diffusion and ensure structural stability during RL exploration. Unlike SMILES (Weininger, 1988; Krenn et al., 2020), SAFE's fragment-based representation imposes strong chemical priors, preventing local invalidity even under aggressive optimization. We further initialize our framework with the pre-trained GenMol (Lee et al., 2025) backbone to inherit learned chemical distributions, allowing the model to focus exclusively on condition alignment rather than learning basic validity.

4 Methodology

As illustrated in Figure 1, we present a unified framework CAGenMol for goal-directed molecular generation that synergizes condition-aware discrete diffusion with reinforcement learning. The

framework is designed to explicitly align molecular generation with complex biochemical objectives beyond pure data distribution modeling.

4.1 Model Architecture.

The core design philosophy of CAGenMol is to bridge the modality gap between heterogeneous biological constraints and discrete chemical space. To achieve this, we structurally decouple constraint perception from molecular reasoning. The architecture is composed of two synergistic modules: a **Unified Constraint Adaptor (UCA)** that projects diverse signals (e.g., 3D geometric pockets or 1D property vectors) into a shared latent semantic space, and a **Condition-Aware Diffusion Backbone** that utilizes these unified representations to bias the discrete denoising trajectory. This design allows a single generative framework to flexibly adapt to both extrinsic structural environments and intrinsic property requirements.

4.1.1 Unified Constraint Adaptor

To map heterogeneous constraint signals into the shared latent space of the diffusion backbone with dimension D, UCA acts as a learnable interface that translates biological and chemical constraints into a unified latent guidance representation.

Extrinsic Constraint: Structure Adaptation. A protein pocket defines the external physicochemical environment that constrains the feasible geometric space and interaction patterns of a ligand (Koshland Jr, 1958; Schneider and Fechner, 2005). To encode this extrinsic constraint, we propose a dual-stream encoding strategy to bridge the gap between implicit evolutionary semantics and explicit surface chemistry. While protein language models capture long-range dependencies, they often lack the granularity required for precise interaction matching. Therefore, we augment semantic features with explicit physicochemical descriptors:

(1) Semantic Stream. We extract residue-level embeddings $\mathbf{H}_{esm} \in \mathbb{R}^{L_{pocket} \times 1280}$ using the pre-trained ESM-2 (Lin et al., 2023), leveraging its evolutionary knowledge to characterize the pocket's biological context.

(2) Physicochemical Stream. To explicitly guide interaction matching, we compute a 5-dimensional feature vector $\mathbf{h}_{phys}$ for each residue (e.g., charge, hydropathy, H-bond potential). These are stacked to form $\mathbf{H}_{phys} \in \mathbb{R}^{L_{pocket} \times 5}$, ensuring the model attends to key binding determinants.

Both streams are projected into the shared model dimension D via independent MLPs to align their semantic manifolds:

$$\tilde{\mathbf{H}}_{esm} = \mathrm{MLP}_{esm}(\mathbf{H}_{esm}^{pocket}) \in \mathbb{R}^{L_{pocket} \times D}, \quad (3)$$

$$\tilde{\mathbf{H}}_{phys} = \mathrm{MLP}_{phys}(\mathbf{H}_{phys}^{pocket}) \in \mathbb{R}^{L_{pocket} \times D}. \quad (4)$$

The projected features are then fused by element-wise summation to yield the unified pocket representation $\mathbf{H}_{fused} = \tilde{\mathbf{H}}_{esm} + \tilde{\mathbf{H}}_{phys}$.

To identify key binding residues without relying on explicit 3D coordinates, we employ a Linear Attention Pooling mechanism. By computing a learnable importance score for each residue, the model autonomously learns to focus on functional hotspots in the pocket. Specifically, the attention weights $\boldsymbol{\alpha} \in \mathbb{R}^{L_{pocket} \times 1}$ are computed as:

$$\boldsymbol{\alpha} = softmax(\mathrm{MLP}_{attn}(\mathbf{H}_{fused})). \quad (5)$$

The final extrinsic condition token is then obtained as a capability-weighted sum, ensuring that the guidance signal is dominated by the most pharmacologically relevant residues:

$$\mathbf{h}_{ext} = \textstyle\sum_{i=1}^{L_{pocket}} \alpha_i \mathbf{H}_{fused}^{(i)} \in \mathbb{R}^{1 \times D}. \quad (6)$$

Intrinsic Constraint: Property Adaptation. Beyond structural fit, drug candidates must satisfy intrinsic property constraints $\mathbf{y} \in \mathbb{R}^K$ (e.g., ADMET). To translate these scalar values into high-dimensional guidance signals compatible with the diffusion sequence, the UCA projects $\mathbf{y}$ into the latent space via a learnable mapping MLP_{int}, producing a property-conditioned token $\mathbf{h}_{int} \in \mathbb{R}^{1 \times D}$. This enables the diffusion model to interpret numerical properties as semantic prompts.

4.2 Condition-Aware Diffusion Backbone

We adapt the BERT-based GenMol (Lee et al., 2025) architecture for conditional generation. Instead of introducing heavy cross-attention modules which require training from scratch, we propose a parameter-efficient **Prompt-based Conditional Denoising** strategy. This approach treats the condition vector not merely as an input, but as a semantic prompt that prefixes the molecular sequence.

Formally, we construct the input embedding $\mathbf{H}_t$ at time step t by prepending the condition token $\mathbf{h}_\mathbf{c}$ (derived from UCA) to the noisy molecular embeddings:

$$\mathbf{H}_t = \left[\mathbf{h}_\mathbf{c}, \mathrm{Embed}(\boldsymbol{z}_t^1), \ldots, \mathrm{Embed}(\boldsymbol{z}_t^L)\right]. \quad (7)$$

This design leverages the **global broadcasting capability** of bidirectional self-attention. Since $\mathbf{h_c}$ is visible to every molecular token at every layer, it serves as a **persistent semantic anchor** throughout the diffusion process. Even when the molecular sequence $\boldsymbol{z}_t$ is heavily corrupted by mask tokens (in the forward process), the unmasked $\mathbf{h_c}$ provides a stable reference signal. This allows the model to effectively bias the denoising distribution $p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{z}_t, \mathbf{h_c})$ toward the target chemical manifold without disrupting the pre-trained structural priors.

4.3 Training and Inference Pipeline.

CAGenMol is optimized and deployed following a three-stage paradigm. First, the model is trained via supervised learning with a discrete diffusion objective, which provides a stable initialization for subsequent optimization. Second, we introduce a step-wise Proximal Policy Optimization (Step-PPO) algorithm to further steer the generation process toward task-specific objectives. Finally, during inference, an Evolutionary Fragment Optimization (EFO) procedure is applied to iteratively refine and improve the generated molecular candidates.

4.3.1 Supervised Learning

We first train CAGenMol via supervised learning to adapt the unconditional backbone to conditioning signals. Following MDLM (Sahoo et al., 2024), we optimize a continuous-time approximation of the negative evidence lower bound (NELBO), which acts as a time-weighted masked language modeling objective over the molecular tokens (details in Appendix E). This stage establishes a stable, condition-aware initialization for subsequent RL alignment.

4.3.2 Step-wise PPO for Diffusion Language Model

While the supervised stage establishes a chemical prior, it prioritizes likelihood over functional desirability, often failing to explore high-reward regions defined by non-differentiable oracles (e.g., docking). Thus, we propose **Step-wise Proximal Policy Optimization (Step-PPO)**. Unlike traditional trajectory-level RL, Step-PPO reformulates discrete diffusion as a fine-grained Markov Decision Process (MDP), enabling precise alignment with complex objectives while preserving generative coherence.

Algorithm. We can interpret the reverse diffusion process as a MDP, which naturally arises from the iterative denoising formulation. At each diffusion time step t, the state $s_t = \boldsymbol{z}_t$ corresponds to the partially masked molecular sequence, and the policy π_θ (parameterized by the diffusion model) defines a distribution over actions $a_t \sim \pi_\theta(\cdot \mid s_t)$, where an action a_t corresponds to selecting categorical tokens to replace the masked positions in $\boldsymbol{z}_t$ during the reverse transition from t to $t-1$. This formulation satisfies the Markov property, as each denoising step depends only on the current sequence state.

Unlike prior diffusion-based RL approaches that treat the entire denoising trajectory as a single action (Zhao et al., 2025; Shao et al., 2024; Yang et al., 2025), we apply policy optimization (Schulman et al., 2017) at each diffusion step. As intermediate masked states lack defined chemical properties, we formulate the task as a **sparse reward problem**, maximizing the terminal reward R evaluated solely at $t = 0$.

To stabilize policy updates, we adopt the clipped surrogate objective from PPO (Schulman et al., 2017). For a specific diffusion step t involving an action a_t, the loss function is defined as:

$$\mathcal{L}_{\text{step}}^{(t)}(\theta) = -\mathbb{E}_{\pi_{\theta_{\text{old}}}}\left[\min\left(r_t(\theta)\hat{A}_t,\ \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right] \tag{8}$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$ denotes the probability ratio between the current and behavioral policies.

To efficiently estimate the signal without an additional value network, we compute the advantage $\hat{A}_t$ using batch-level reward statistics:

$$\hat{A}_t = \frac{R - \mu_{\mathcal{B}}}{\sigma_{\mathcal{B}} + \epsilon}, \tag{9}$$

where R represents the terminal reward of the current trajectory, while $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}$ denote the mean and standard deviation of valid rewards within the sampling batch $\mathcal{B}$.

Moreover, rewards are only well-defined for chemically valid molecules. So we introduce a Validity Mask $\mathcal{M}_{valid} \in \{0, 1\}$, which activates policy updates only for valid trajectories and effective denoising steps. To ensure stability, we introduce a mask $\mathcal{M}_{valid} \in \{0, 1\}$ to restrict updates to chemically valid trajectories. Given the pre-trained backbone's high validity rate ($> 90\%$), this mechanism imposes negligible overhead and serves effectively as a noise filter, preventing undefined reward signals from corrupting gradient estimation.

The final optimization objective over a batch of size B is given by:

$$\mathcal{L}_{\text{batch}} = \frac{\sum_{b=1}^{B} \mathcal{M}_{valid}^{(b)} \left(\sum_t \mathcal{L}_{\text{step}}^{(b,t)} + \beta \mathcal{L}_{\text{ent}}^{(b,t)} \right)}{\sum_{b=1}^{B} \mathcal{M}_{valid}^{(b)}}, \quad (10)$$

where $\mathcal{L}_{\text{ent}}$ denotes the entropy regularization term. This design ensures that policy updates are driven exclusively by valid molecular trajectories that satisfy the imposed conditions.

Reward Design. While the optimization algorithm is generic, the reward function encodes task-specific objectives. We design task-specific terminal reward functions for structure-conditioned and property-conditioned molecular generation, both evaluated on the fully denoised molecule at the final step $t = 0$.

For structure-conditioned generation, the primary objective is to optimize the protein-ligand interaction strength. We quantify this using the standard Vina score (S_{dock}) computed by AutoDock Vina (Eberhardt et al., 2021), which approximates the Gibbs free energy of binding (ΔG) in kcal/mol. Since raw docking scores are unbounded and highly target-dependent, we introduce a reference affinity threshold S_{ref} (determined empirically from the reference ligand's affinity, see Sec. 5.1). We then define the affinity margin $\Delta = S_{\text{ref}} - S_{\text{dock}}$ to measure relative improvement. The structure-based reward is defined as:

$$R_{\text{struct}} = \text{sign}(\Delta) \cdot \Delta^2 + \lambda_1 \cdot \text{QED} + \lambda_2 \cdot \text{SA}, \quad (11)$$

where QED (Bickerton et al., 2012) and SA (Ertl and Schuffenhauer, 2009) encourage drug-likeness and synthetic accessibility. λ_1 and λ_2 are balancing coefficients used to normalize the scales of different objectives. This quadratic formulation implicitly prioritizes the refinement of pharmaceutical properties when the binding affinity approaches the reference threshold.

For property-conditioned generation, let $\mathbf{y}_{\text{tgt}} \in \mathbb{R}^K$ denote the target property vector and $\hat{\mathbf{y}} \in \mathbb{R}^K$ the predicted properties. To harmonize heterogeneous property scales into a bounded reward space $[0, 1]$, we map the prediction error to a similarity score using a weighted Gaussian kernel:

$$R_{\text{prop}} = \sum_{k=1}^{K} \omega_k \exp\left(-\frac{(\hat{y}_k - y_{\text{tgt},k})^2}{2\sigma_k^2} \right). \quad (12)$$

To calibrate for heterogeneous scales and optimization difficulty, we derive σ_k and ω_k from statistics of 1,000 molecules sampled from the initial policy π_{init}. Specifically, σ_k is defined as the empirical standard deviation $\text{Std}_{\mathcal{D}}(y_k)$ of these samples to normalize diverse numerical ranges. Meanwhile, the weight $\omega_k = \varepsilon_k / \sum_{j=1}^{K} \varepsilon_j$ is assigned proportional to the mean absolute error $\varepsilon_k = \mathbb{E}_{\mathcal{D}}[|\hat{y}_k^{(0)} - y_k|]$ on this batch, adaptively prioritizing properties that are initially harder to satisfy.

4.3.3 Evolutionary Fragment Optimization

To mitigate sampling stochasticity, we introduce **Evolutionary Fragment Optimization (EFO)** to perform gradient-free hill-climbing at inference time. EFO iteratively refines candidates by resampling masked substructures via the conditional diffusion backbone. Formally, for a molecule x, we apply a mask to select $x_{\text{mask}} \subset x$ and sample a new candidate x' conditioned on the remaining context:

$$x' \sim p_\theta(x_{\text{mask}} \mid x \setminus x_{\text{mask}}, \mathbf{h_c}). \quad (13)$$

The fragment vocabulary $\mathcal{V}$ is dynamically updated by decomposing generated candidates and retaining the top-K structures based on property scores $S(\cdot)$:

$$\mathcal{V}_{t+1} \leftarrow \text{TopK}\big(\mathcal{V}_t \cup \text{Decompose}(x'),\ S\big). \quad (14)$$

This loop concentrates the search on high-value chemical regions (see Algorithm 1 in Appendix).

5 Experiments

We conduct three sets of experiments to evaluate CAGenMol under (i) structure-conditioned generation, (ii) multi-target property-conditioned generation and (iii) dual-conditioned generation. All training and evaluation are performed on a single **NVIDIA A800** GPU.

5.1 Structure-Conditioned Generation

This task aims to generate small-molecule ligands that bind favorably to a given protein pocket under fixed receptor geometry.

Data. Following standard practice in pocket-aware molecular generation (Peng et al., 2022; Guan et al., 2023), we use the CrossDocked2020 dataset (Francoeur et al., 2020). The processed dataset contains approximately 100,000 protein pocket–ligand complexes for training, with 100 target protein pockets held out for evaluation.

Methods	Vina Dock (↓)	High Affinity (↑)	QED (↑)	SA (↑)	Diversity (↑)	Success Rate (↑)
Reference Set	-7.45	-	0.48	0.73	-	25.0%
TargetDiff	-7.80	58.1%	0.48	0.58	0.72	10.5%
FLAG	-5.63	-	0.49	0.70	0.70	14.1%
Pocket2Mol	-7.15	48.4%	0.56	0.74	0.69	24.4%
DecompDiff	-8.39	64.4%	0.45	0.61	0.68	24.5%
MolCRAFT	**-9.25**	59.1%	0.46	0.62	0.61	36.1%
RGA + Vina	-8.01	64.4%	<u>0.57</u>	0.71	0.41	46.2%
DecompOpt	<u>-8.98</u>	73.5%	0.48	0.65	0.60	52.5%
MOLCHORD	-8.59	74.6%	0.56	<u>0.78</u>	0.71	53.4%
CAGenMol	-8.41	**82.3%**	**0.70**	**0.89**	**0.75**	**69.7%**

Table 1: Comparison on the CrossDocked2020 benchmark. We report the average metrics across 100 test pockets. The best results are highlighted in **bold**, and the second best are <u>underlined</u>.

Variant	**Vina Dock** (↓)	**QED** (↑)	**SA** (↑)	**Diversity** (↑)	**Success Rate** (↑)
SFT (Base)	-6.47	0.53	0.77	**0.88**	14.3%
SFT (w/o Attn.)	-6.60	0.55	0.76	0.86	17.5%
SFT (w/o Phys.)	-6.55	0.54	0.78	0.87	15.8%
SFT (Full UCA)	-6.61	0.58	0.77	0.86	19.2%
Full Model (Step-PPO)	**-8.41**	**0.70**	**0.89**	0.80	**69.7%**

Table 2: Ablation study of UCA components and RL training. All variants except the final one are trained solely with Supervised Fine-Tuning (SFT).

Baselines. We compare CAGenMol against representative structure-conditioned generation and optimization methods evaluated under the same protocol, including TargetDiff (Guan et al., 2023), Pocket2Mol (Peng et al., 2022), DecompDiff (Guan et al., 2024), MolCRAFT (Qu et al., 2024), RGA+Vina (Fu et al., 2022), DecompOpt (Zhou et al., 2024), and MOLCHORD (Zhang et al., 2025).

Training and Inference Protocol. We adopt a two-stage training strategy: supervised fine-tuning followed by Step-PPO maximization of Eq. 11. We set $S_{\text{ref}} = -9.0$, where the quadratic term implicitly shifts optimization focus to QED and SA as affinity improves. To balance the magnitude disparity between Vina scores and property metrics, we set $\lambda_1 = 7/3$ and $\lambda_2 = 5/6$. EFO is excluded in this benchmark to ensure fair comparison. See Appendix D for details.

Evaluation Metrics. We evaluate 100 molecules generated for each pocket using the following metrics: (1) **Vina Dock**: Binding affinity calculated by AutoDock Vina (Eberhardt et al., 2021) under the protocol of (Guan et al., 2023); (2) **QED**: Quantitative Estimate of Drug-likeness (Bickerton et al., 2012); (3) **SA**: Synthetic Accessibility score (Ertl and Schuffenhauer, 2009); and (4) **Diversity**: The average pairwise Tanimoto distance among generated molecules per pocket. Following prior benchmarks (Long et al., 2022; Guan et al., 2024), we also report the **Success Rate**, defined as the percentage of valid molecules simultaneously satisfying $\text{Vina} < -8.18$, $\text{QED} > 0.25$, and $\text{SA} > 0.59$. (5) **High Affinity**: Following prior evaluation protocols, we compute High Affinity as the percentage of generated molecules whose docking scores are no worse than those of the test-set ligands.

Results. Table 1 shows that CAGenMol establishes a new state-of-the-art with a **69.7% Success Rate**, surpassing the best baseline by over 16%. Unlike methods that sacrifice molecular quality for raw docking scores, CAGenMol achieves a superior balance, dominating in **QED** and **SA** while maintaining strong affinity. Notably, it also retains the highest **Diversity**, demonstrating that Step-PPO effectively optimizes binding without suffering from the mode collapse typically associated with RL. (See Appendix A for visual examples of generated molecules). We also conducted experiments on an additional benchmark following (Zheng et al., 2024) to further demonstrate robustness (see Appendix H).

Model	1IEP	3EML	3NY8	4RLU	4UNN	5MO4	7L11	Avg
3DSBDD	-9.05±0.38	-10.02±0.15	-10.10±0.24	-9.80±0.55	-8.23±0.30	-8.71±0.45	-8.47±0.18	-9.20
AutoGrow4	-13.23±0.11	-13.03±0.09	-11.70±0.00	-11.20±0.00	-11.14±0.12	-10.38±0.27	-8.84±0.33	-11.36
Pocket2Mol	-10.17±0.53	-12.25±0.27	-11.89±0.16	-10.57±0.12	-12.20±0.34	-10.07±0.62	-9.74±0.38	-10.98
PocketFlow	-12.49±0.70	-9.25±0.29	-8.56±0.35	-9.65±0.25	-7.90±0.78	-7.80±0.42	-8.35±0.31	-9.14
ResGen	-10.97±0.29	-9.25±0.95	-10.96±0.42	-11.75±0.42	-9.41±0.23	-10.34±0.39	-8.74±0.24	-10.20
DST	-10.95±0.57	-10.67±0.24	-10.54±0.22	-10.88±0.37	-9.71±0.19	-10.03±0.36	-8.33±0.41	-10.16
GraphGA	-10.03±0.41	-9.89±0.25	-9.94±0.15	-10.22±0.39	-9.32±0.51	-9.29±0.20	-7.75±0.32	-9.49
MIMOSA	-10.96±0.57	-10.69±0.24	-10.51±0.23	-10.81±0.39	-9.66±0.25	-10.02±0.36	-8.33±0.41	-10.14
MolDQN	-6.73±0.12	-6.51±0.15	-7.09±0.16	-6.79±0.26	-5.92±0.26	-6.27±0.10	-6.87±0.20	-6.60
Pasithea	-10.86±0.29	-10.31±0.09	-10.69±0.27	-10.92±0.35	-9.69±0.32	-9.77±0.21	-8.06±0.22	-10.04
REINVENT	-9.87±0.31	-9.48±0.39	-9.61±0.36	-9.69±0.29	-8.70±0.25	-8.92±0.38	-7.25±0.21	-9.07
SCREENING	-10.86±0.26	-10.90±0.54	-10.73±0.45	-10.86±0.22	-9.80±0.23	-9.91±0.30	-8.15±0.26	-10.17
SELFIES-VAE-BO	-10.15±0.60	-9.76±0.12	-9.99±0.28	-10.00±0.23	-9.02±0.33	-9.18±0.39	-7.75±0.22	-9.41
SMILES GA	-9.56±0.17	-9.56±0.37	-10.00±0.26	-9.61±0.19	-8.80±0.20	-9.21±0.23	-7.54±0.32	-9.18
SMILES LSTM HC	-10.38±0.21	-10.30±0.15	-10.19±0.12	-10.49±0.49	-9.36±0.17	-9.71±0.43	-7.90±0.26	-9.76
SMILES-VAE-BO	-9.93±0.22	-9.78±0.10	-9.96±0.29	-10.05±0.20	-9.03±0.30	-9.18±0.39	-7.74±0.25	-9.38
CAGenMol	-12.33±0.11	-12.26±0.18	-11.90±0.32	-12.40±0.22	-11.85±0.19	-10.86±0.22	-8.97±0.25	-11.51
CAGenMol + EFO	-12.83±0.26	-12.76±0.22	-12.31±0.19	-12.49±0.28	-11.88±0.20	-11.08±0.24	-9.07±0.31	**-11.77**

Table 3: Top-10 average docking scores on additional benchmark(lower is better).

Setting	Property	De novo	Step-PPO	Step-PPO + EFO	Target
Setting 1	HIA	0.97	0.98	0.99	↑ (1)
	BBB	0.53	0.89	0.89	↑ (1)
	Ames	0.37	0.21	0.15	↓ (0)
Setting 2	CYP3A4sub	0.48	0.64	0.69	↑ (1)
	LogP	2.3	3.4	3.5	[3,5]
	DILI	0.55	0.30	0.29	↓ (0)
Setting 3	Solubility	-3.7	-1.2	-1.1	> -1
	hERG	0.60	0.35	0.32	↓ (0)
	Pgp_Sub	0.22	0.16	0.12	↓ (0)

Table 4: ADMET evaluation results under different settings.

5.2 Property-Conditioned Generation

We evaluate intrinsic property conditioning on three practically motivated ADMET settings. We use MiniMol (Kläser et al., 2024) as the property predictor to provide supervision and reward signals.

ADMET settings. We consider three multi-constraint targets: **Setting 1 (CNS drugs):** $\mathrm{HIA} = 1$, $\mathrm{BBB} = 1$, $\mathrm{Ames} = 0$. **Setting 2 (Hepatically metabolized drugs):** $\mathrm{CYP3A4_{sub}} = 1$, $\mathrm{DILI} = 0$, $\mathrm{LogP} \in [3, 5]$. **Setting 3 (Peripheral drugs):** $\mathrm{Solubility} > -1$, $\mathrm{hERG} = 0$, $\mathrm{Pgp\_Sub} = 0$.

Training and Inference. For each setting, we first sample 10,000 *de novo* generated molecules from the unconditional base model, and annotate each molecule using MiniMol. For binary properties, we convert predicted probabilities into hard labels $\{0, 1\}$ to form a pseudo-labeled training set, and then train CAGenMol with supervised learning under the corresponding target property vector.

And then, we further optimize CAGenMol using Step-PPO where the terminal reward is computed from MiniMol-predicted probabilities (without hard-thresholding) to preserve gradient-free but smooth optimization signals. Finally, due to computational constraints, we run EFO for three generations to refine the candidate pool. Detailed hyperparameters and computational costs are provided in Appendix D.

Results. We visualize the distribution shift of each target property via histograms in Figure 2 across three distinct stages: (i) Unconditional Generation from the base model; (ii) Step-PPO Optimized; and (iii) Step-PPO + EFO, where the candidates are further refined during inference. This progression highlights how Step-PPO effectively shifts the molecular distribution toward the desired properties, while EFO provides a final sharpening of constraint satisfaction.

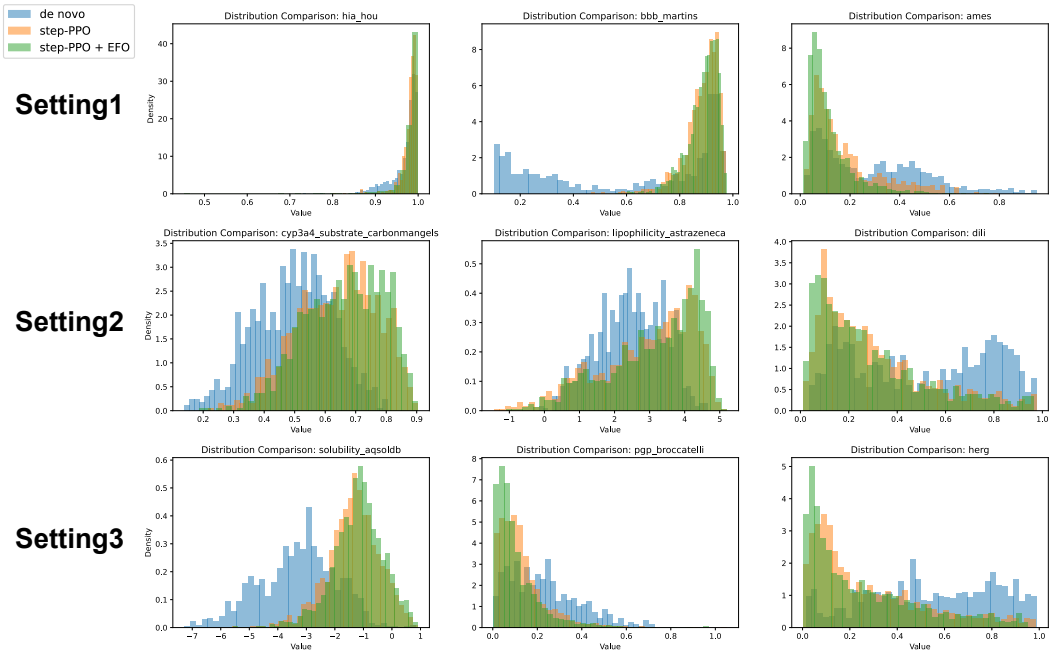

Figure 2: Histograms of Distribution shift under ADMET constraints for three settings.

5.3 Dual-Conditioned Generation

In real-world drug discovery, candidate molecules are required to simultaneously achieve strong binding affinity to the target protein and favorable ADMET properties. To evaluate the capability of CAGenMol under such realistic constraints, we conduct a dual-conditioned generation experiment that jointly enforces structure-based binding and toxicity-related property requirements.

We consider the protein 3o96_A, an important therapeutic target in the PI3K–AKT signaling pathway. For this target, we condition molecule generation on both the fixed protein pocket and a safety constraint requiring Ames-negative predictions. Using the same training and optimization pipeline as in previous experiments, we generate 100 candidate molecules and evaluate them in terms of docking affinity and predicted Ames toxicity.

Results in Table 5 show that CAGenMol achieves the best overall performance under both criteria, producing molecules with superior docking scores while maintaining the highest proportion of Ames-negative candidates among all compared methods. This demonstrates that CAGenMol can effectively balance binding optimization and toxicity avoidance within a unified framework, highlighting its practical value for realistic drug design scenarios. Detailed experimental settings and additional analyzes are provided in Appendix G.

Methods	Vina Dock (↓)	QED (↑)	SA (↑)	Ames (↓)
Pocket2Mol	-10.38	0.71	0.70	0.54
TargetDiff	-10.80	0.39	0.51	0.44
FLAG	-6.38	0.60	0.67	0.36
MolCRAFT	-11.33	0.43	0.66	0.49
DecompDiff	**-12.33**	0.26	0.54	0.57
w/o Ames Condition	-10.13	0.83	**0.88**	0.34
CAGenMol	-9.94	**0.84**	**0.88**	**0.18**

Table 5: Comparison under dual-conditioned generation on the 3o96_A pocket.

6 Conclusion

We propose CAGenMol, a unified framework for goal-directed molecular generation. By effectively handling structural, property, and dual constraints, CAGenMol achieves state-of-the-art performance across diverse benchmarks while maintaining quality and diversity, demonstrating strong potential for practical drug discovery applications.

Limitations

Despite the improvements of CAGenMol, several limitations remain. Limited computational resources prevented large-scale training and extensive benchmarking, forcing us to initialize the model with pretrained weights rather than training from scratch. Additionally, the reliance on predictions from tools and models like ESM-2, AutoDock Vina and MiniMol, without accounting for potential prediction errors, may impact overall performance.

References

Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. 2021. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems*, 34:17981–17993.

Viraj Bagal, Rishal Aggarwal, PK Vinod, and U Deva Priyakumar. 2021. Molgpt: molecular generation using a transformer-decoder model. *Journal of chemical information and modeling*, 62(9):2064–2076.

G Richard Bickerton, Gaia V Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L Hopkins. 2012. Quantifying the chemical beauty of drugs. *Nature chemistry*, 4(2):90–98.

Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. 2016. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47:20–33.

Jerome Eberhardt, Diogo Santos-Martins, Andreas F Tillack, and Stefano Forli. 2021. Autodock vina 1.2. 0: new docking methods, expanded force field, and python bindings. *Journal of chemical information and modeling*, 61(8):3891–3898.

Peter Ertl and Ansgar Schuffenhauer. 2009. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of cheminformatics*, 1:1–11.

Wei Feng, Lvwei Wang, Zaiyun Lin, Yanhao Zhu, Han Wang, Jianqiang Dong, Rong Bai, Huting Wang, Jielong Zhou, Wei Peng, and 1 others. 2024. Generation of 3d molecules in pockets via a language model. *Nature Machine Intelligence*, 6(1):62–73.

Leonardo LG Ferreira and Adriano D Andricopulo. 2019. Admet modeling approaches in drug discovery. *Drug discovery today*, 24(5):1157–1165.

Paul G Francoeur, Tomohide Masuda, Jocelyn Sunseri, Andrew Jia, Richard B Iovanisci, Ian Snyder, and David R Koes. 2020. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of chemical information and modeling*, 60(9):4200–4215.

Cong Fu, Xiner Li, Blake Olson, Heng Ji, and Shuiwang Ji. 2024. Fragment and geometry aware tokenization of molecules for structure-based drug design using language models. *arXiv preprint arXiv:2408.09730*.

Tianfan Fu, Wenhao Gao, Connor Coley, and Jimeng Sun. 2022. Reinforced genetic algorithm for structure-based drug design. *Advances in Neural Information Processing Systems*, 35:12325–12338.

Tianfan Fu, Wenhao Gao, Cao Xiao, Jacob Yasonik, Connor W Coley, and Jimeng Sun. 2021. Differentiable scaffolding tree for molecular optimization. *arXiv preprint arXiv:2109.10469*.

Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 2023. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543*.

Jiaqi Guan, Xiangxin Zhou, Yuwei Yang, Yu Bao, Jian Peng, Jianzhu Ma, Qiang Liu, Liang Wang, and Quanquan Gu. 2024. Decompdiff: diffusion models with decomposed priors for structure-based drug design. *arXiv preprint arXiv:2403.07902*.

Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf H Roohani, Jure Leskovec, Connor W Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. 2021. Therapeutics data commons: Machine learning datasets and tasks for drug discovery and development. In *NeurIPS Track Datasets and Benchmarks*.

James P Hughes, Stephen Rees, S Barrett Kalindjian, and Karen L Philpott. 2011. Principles of early drug discovery. *British journal of pharmacology*, 162(6):1239–1249.

Moksh Jain, Sharath Chandra Raparthy, Alex Hernández-Garcıa, Jarrid Rector-Brooks, Yoshua Bengio, Santiago Miret, and Emmanuel Bengio. 2023. Multi-objective gflownets. In *International conference on machine learning*, pages 14631–14653. PMLR.

Jan H Jensen. 2019. A graph-based genetic algorithm and generative model/monte carlo tree search for the exploration of chemical space. *Chemical science*, 10(12):3567–3572.

Kerstin Kläser, Błażej Banaszewski, Samuel Maddrell-Mander, Callum McLean, Luis Müller, Ali Parviz, Shenyang Huang, and Andrew Fitzgibbon. 2024. Minimol: A parameter-efficient foundation model for molecular learning. *arXiv preprint arXiv:2404.14986*.

Daniel E Koshland Jr. 1958. Application of a theory of enzyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences*, 44(2):98–104.

Mario Krenn, Florian Häse, AkshatKumar Nigam, Pascal Friederich, and Alan Aspuru-Guzik. 2020. Self-referencing embedded strings (selfies): A 100% robust molecular string representation. *Machine Learning: Science and Technology*, 1(4):045024.

Seul Lee, Karsten Kreis, Srimukh Prasad Veccham, Meng Liu, Danny Reidenbach, Yuxing Peng, Saee Paliwal, Weili Nie, and Arash Vahdat. 2025. Genmol: A drug discovery generalist with discrete diffusion. *arXiv preprint arXiv:2501.06158*.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, and 1 others. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130.

Meng Liu, Youzhi Luo, Kanji Uchino, Koji Maruhashi, and Shuiwang Ji. 2022. Generating 3d molecules for target protein binding. *arXiv preprint arXiv:2204.09410*.

Hannes H Loeffler, Jiazhen He, Alessandro Tibo, Jon Paul Janet, Alexey Voronov, Lewis H Mervin, and Ola Engkvist. 2024. Reinvent 4: modern ai–driven generative molecule design. *Journal of Cheminformatics*, 16(1):20.

Siyu Long, Yi Zhou, Xinyu Dai, and Hao Zhou. 2022. Zero-shot 3d drug design by sketching and generating. *Advances in Neural Information Processing Systems*, 35:23894–23907.

Aaron Lou, Chenlin Meng, and Stefano Ermon. 2024. Discrete diffusion language modeling by estimating the ratios of the data distribution. *International Conference on Machine Learning*.

Shitong Luo, Jiaqi Guan, Jianzhu Ma, and Jian Peng. 2021. A 3d generative model for structure-based drug design. *Advances in Neural Information Processing Systems*, 34:6229–6239.

Emmanuel Noutahi, Cristian Gabellini, Michael Craig, Jonathan SC Lim, and Prudencio Tossou. 2024. Gotta be safe: a new framework for molecular design. *Digital Discovery*, 3(4):796–804.

Xingang Peng, Shitong Luo, Jiaqi Guan, Qi Xie, Jian Peng, and Jianzhu Ma. 2022. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In *International conference on machine learning*, pages 17644–17655. PMLR.

Pavel G Polishchuk, Timur I Madzhidov, and Alexandre Varnek. 2013. Estimation of the size of drug-like chemical space based on gdb-17 data. *Journal of computer-aided molecular design*, 27(8):675–679.

Yanru Qu, Keyue Qiu, Yuxuan Song, Jingjing Gong, Jiawei Han, Mingyue Zheng, Hao Zhou, and Wei-Ying Ma. 2024. Molcraft: structure-based drug design in continuous parameter space. *arXiv preprint arXiv:2404.12141*.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.

Philipp Renz, Dries Van Rompaey, Jörg Kurt Wegner, Sepp Hochreiter, and Günter Klambauer. 2019. On failure modes in molecule generation and optimization. *Drug Discovery Today: Technologies*, 32:55–63.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T Chiu, Alexander Rush, and Volodymyr Kuleshov. 2024. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*.

Gisbert Schneider and Uli Fechner. 2005. Computer-based de novo design of drug-like molecules. *Nature reviews Drug discovery*, 4(8):649–663.

Arne Schneuing, Charles Harris, Yuanqi Du, Kieran Didi, Arian Jamasb, Ilia Igashov, Weitao Du, Carla Gomes, Tom L Blundell, Pietro Lio, and 1 others. 2024. Structure-based drug design with equivariant diffusion models. *Nature Computational Science*, 4(12):899–909.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis K Titsias. 2024. Simplified and generalized masked diffusion for discrete data. *Advances in neural information processing systems*.

Jacob O Spiegel and Jacob D Durrant. 2020. Autogrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *Journal of cheminformatics*, 12(1):25.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Haorui Wang, Marta Skreta, Cher-Tian Ser, Wenhao Gao, Lingkai Kong, Felix Strieth-Kalthoff, Chenru Duan, Yuchen Zhuang, Yue Yu, Yanqiao Zhu, and 1 others. 2024. Efficient evolutionary search over chemical space with large language models. *arXiv preprint arXiv:2406.16976*.

David Weininger. 1988. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36.

Jingyi Yang, Guanxu Chen, Xuhao Hu, and Jing Shao. 2025. Taming masked diffusion language models via consistency trajectory reinforcement learning with fewer decoding step. *arXiv preprint arXiv:2509.23924*.

Naruki Yoshikawa, Kei Terayama, Masato Sumita, Teruki Homma, Kenta Oono, and Koji Tsuda. 2018. Population-based de novo molecule generation, using grammatical evolution. *Chemistry Letters*, 47(11):1431–1434.

Zebin You, Shen Nie, Xiaolu Zhang, Jun Hu, Jun Zhou, Zhiwu Lu, Ji-Rong Wen, and Chongxuan Li. 2025. Llada-v: Large language diffusion models with visual instruction tuning. *arXiv preprint arXiv:2505.16933*.

Wei Zhang, Zekun Guo, Yingce Xia, Peiran Jin, Shufang Xie, Tao Qin, and Xiang-Yang Li. 2025. Molchord: Structure-sequence alignment for protein-guided drug design. *arXiv preprint arXiv:2510.27671*.

Zaixi Zhang and Qi Liu. 2023. Learning subpocket prototypes for generalizable structure-based drug design. In *International Conference on Machine Learning*, pages 41382–41398. PMLR.

Zaixi Zhang, Yaosen Min, Shuxin Zheng, and Qi Liu. 2023. Molecule generation for target protein binding with structural motifs. In *The eleventh international conference on learning representations*.

Siyan Zhao, Devaansh Gupta, Qinqing Zheng, and Aditya Grover. 2025. d1: Scaling reasoning in diffusion large language models via reinforcement learning. *arXiv preprint arXiv:2504.12216*.

Kangyu Zheng, Yingzhou Lu, Zaixi Zhang, Zhongwei Wan, Yao Ma, Marinka Zitnik, and Tianfan Fu. 2024. Structure-based drug design benchmark: do 3d methods really dominate? *arXiv preprint arXiv:2406.03403*.

Xiangxin Zhou, Xiwei Cheng, Yuwei Yang, Yu Bao, Liang Wang, and Quanquan Gu. 2024. Decompopt: Controllable and decomposed diffusion models for structure-based molecular optimization. *arXiv preprint arXiv:2403.13829*.

Zhenpeng Zhou, Steven Kearnes, Li Li, Richard N Zare, and Patrick Riley. 2019. Optimization of molecules via deep reinforcement learning. *Scientific reports*, 9(1):10752.

A Case Study

Figure 3 visualizes two representative pockets from the benchmark. For each pocket, we show the reference ligand from the dataset and two examples generated by CAGenMol with their binding poses, 2d graph and metrics. Results show that our method generate better molecules in every metric.

B Residue-level Physicochemical Feature Definition

To explicitly encode surface biochemical properties of protein pockets, we construct a residue-level physicochemical feature representation based on amino acid identity. For each residue, we define a five-dimensional feature vector capturing hydrophobicity, electrostatic charge, polarity, and hydrogen-bonding capability. These features are computed deterministically and do not require additional learning.

Hydropathy. We adopt the Kyte–Doolittle hydropathy index for each amino acid, which quantifies residue hydrophobicity on a continuous scale. To ensure numerical stability and compatibility with neural network inputs, the raw hydropathy values are normalized by a factor of 5. For example, isoleucine (I) has a value of 4.5, while arginine (R) has a value of -4.5.

Electrostatic Charge. Residue charge is encoded as a discrete scalar. Positively charged residues arginine (R) and lysine (K) are assigned a value of +1, negatively charged residues aspartic acid (D) and glutamic acid (E) are assigned -1, and histidine (H) is assigned a partial charge of +0.1 to reflect its conditional protonation state. All other residues are assigned 0.

Polarity. Polarity is represented as a binary indicator. Polar residues (R, N, D, Q, E, H, K, S, T, Y) are assigned 1, while nonpolar residues are assigned 0.

Hydrogen Bond Acceptor. A binary feature indicates whether a residue can act as a hydrogen bond acceptor. Residues capable of accepting protons (D, E, N, Q, H, S, T, Y) are assigned 1, and all others are assigned 0.

Hydrogen Bond Donor. Similarly, hydrogen bond donor capability is encoded as a binary indicator. Residues that can donate protons (R, K, W, N, Q, H, S, T, Y) are assigned 1, while the remaining residues are assigned 0.

This deterministic encoding provides an explicit and interpretable description of residue surface chemistry, complementing the learned semantic representations extracted from protein language models.

C Structured Context Segment of Conditions

To enable conditional generation, we augment the input sequence with a structured context segment that encodes the conditioning signal. Rather than injecting a raw latent vector, the context is wrapped with dedicated special tokens to explicitly indicate its semantic boundaries and type.

Concretely, the context segment is constructed as

$$\langle \texttt{boc} \rangle \, \langle \texttt{boe} \rangle \, \mathbf{h}_{ext} \, \langle \texttt{eoe} \rangle \, \langle \texttt{boi} \rangle \, \mathbf{h}_{int} \, \langle \texttt{eoi} \rangle \, \langle \texttt{eoc} \rangle,$$

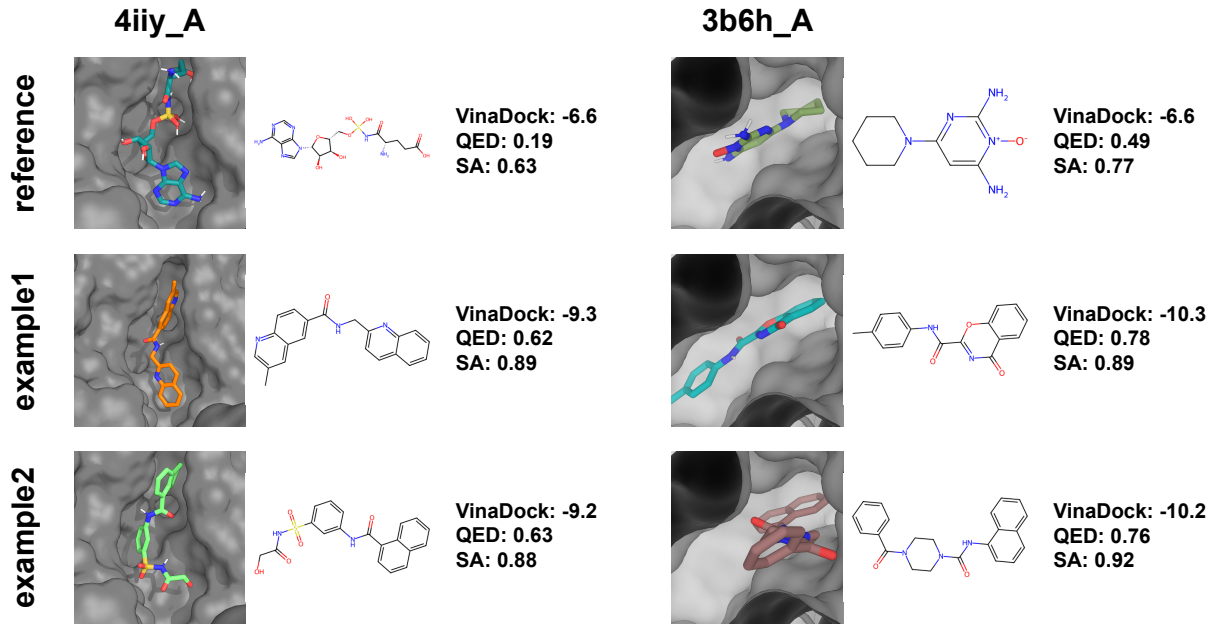

Figure 3: Case Study. Two pockets are shown. For each pocket, we visualize the reference ligand and two CAGenMol-generated ligands with their docking poses in the pocket.

where $\mathbf{h}_{ext}$ and $\mathbf{h}_{int}$ denote the extrinsic (structural) and intrinsic (property) condition embeddings produced by the Unified Constraint Adaptor, respectively. The special tokens ⟨boc⟩/⟨eoc⟩ mark the beginning and end of the entire context segment, while ⟨boe⟩/⟨eoe⟩ and ⟨boi⟩/⟨eoi⟩ delimit the extrinsic and intrinsic condition blocks. All special tokens are associated with learnable word embeddings and are processed identically to molecular tokens within the Transformer.

D Implementation Details

Our training pipeline consists of two phases: supervised fine-tuning and reinforcement learning optimization.

Model Architecture. Our model is based on a BERT-style architecture with around 100 million parameters, comprising 12 Transformer encoder layers. Each layer employs multi-head self-attention and feed-forward sublayers, consistent with the original BERT design.

Supervised Fine-Tuning. We first adapt the unconditional GenMol backbone to the structure-conditioned task. The model is fine-tuned end-to-end on the CrossDocked2020 training set using the **AdamW** optimizer. We set the learning rate to 5×10^{-5} and the batch size to 256. Training is conducted for 10 epochs in precision bf16, which requires approximately 1.5 hours on a single NVIDIA A800 GPU.

Step-PPO Optimization. Starting from the supervised checkpoint, we further optimize the model via Step-PPO to maximize the composite reward defined in Eq. 11.

- **Reward Configuration:** We set the reference affinity barrier to $S_{\text{ref}} = -9$, which is stricter than the evaluation success threshold of -8.18. The QED and SA components are computed using the TDC oracle (Huang et al., 2021).
- **Hyperparameters:** We use a learning rate of 1×10^{-5} and a batch size of 128. The PPO clipping parameter is set to $\epsilon = 0.2$, and the entropy regularization coefficient is set to $\beta = 0.01$. Optimization runs for up to 150 steps, with 2 epochs per step.
- **Sampling Stopping:** During rollouts, we use a sampling temperature of 0.5. We employ an early stopping mechanism that terminates training when the batch success rate exceeds 80%.
- **Compute Cost:** Due to the computational expense of on-policy docking evaluations, this phase requires approximately 190 GPU-hours.

Inference Settings. For the CrossDocked2020 benchmark comparison, we generate molecules using the Step-PPO fine-tuned model without the

Evolutionary Fragment Optimization (EFO) module to strictly adhere to the generation budget of baseline methods.

E Details of Supervised Learning

We first train CAGenMol in a supervised manner to provide a stable initialization for subsequent optimization stages. Here we will continue using the symbols from Section 3.2.

Given a conditioning context $\mathbf{c}$, the denoising network is trained to predict the original molecular tokens from the masked sequence. Since the context $\mathbf{c}$ serves purely as external guidance and is not part of the generation target, no loss is applied to the condition tokens during supervised training. Instead, the objective is defined solely over the molecular sequence.

Following MDLM (Sahoo et al., 2024), we optimize a continuous-time approximation of the negative evidence lower bound (NELBO):

$$\mathcal{L}_{\text{NELBO}} = \mathbb{E}_q \int_0^1 \frac{\alpha_t'}{1-\alpha_t} \sum_{l=1}^{L} \log \left\langle \boldsymbol{x}_\theta^l(\boldsymbol{z}_t, t, \mathbf{c}), \boldsymbol{x}^l \right\rangle \, \mathrm{d}t, \tag{15}$$

where $\boldsymbol{x}_\theta^l(\boldsymbol{z}_t, t, \mathbf{c})$ denotes the predicted categorical distribution for the l-th molecular token conditioned on the noisy sequence $\boldsymbol{z}_t$, diffusion time t, and the context $\mathbf{c}$. This objective corresponds to a time-weighted masked language modeling loss over molecular tokens.

This supervised stage adapts the pre-trained unconditional backbone to operate in the presence of conditioning signals, enabling the model to incorporate contextual information into the denoising process while retaining its original diffusion formulation. It therefore establishes a condition-aware initialization for subsequent optimization.

F Details of Evolutionary Fragment Optimization

Evolutionary Fragment Optimization (EFO) is an iterative evolutionary procedure that integrates fragment-level exploration with masked discrete diffusion. Unlike classical genetic algorithms (Jensen, 2019; Fu et al., 2021)that rely on random atomic mutations, EFO operates on a dynamic vocabulary of chemically meaningful fragments and employs a structured remasking strategy, enabling efficient traversal of chemical space while optimizing task-specific objectives. The optimization process is centered around a fragment vocabulary $\mathcal{V}$, which serves as a genetic pool for molecule construction and evolution.

To prioritize high-value substructures, we define a scoring function $S(f_k)$ for each fragment f_k extracted from a source dataset $\mathcal{D}$. The score reflects the average target property value of molecules containing the fragment:

$$S(f_k) = \frac{1}{|\mathcal{S}(f_k)|} \sum_{x \in \mathcal{S}(f_k)} y(x), \tag{16}$$

where $\mathcal{S}(f_k) = \{x \in \mathcal{D} : f_k \text{ is a subgraph of } x\}$. The initial vocabulary $\mathcal{V}$ is constructed by selecting the top-V fragments ranked by $S(f_k)$ and is dynamically updated throughout the optimization process to incorporate newly discovered high-scoring fragments.

The EFO procedure iterates over a generative cycle consisting of four tightly integrated stages: initialization, mutation, guided reconstruction, and vocabulary evolution.

1. **Initialization via Fragment Attachment.** Each iteration begins by constructing a seed molecule x_{init}. Two fragments are randomly sampled from the current vocabulary $\mathcal{V}$ and attached to form a valid Sequential Attachment-based Fragment Embedding (SAFE) representation. This initialization strategy ensures that the starting molecules already contain substructures statistically correlated with favorable target properties.

2. **Mutation via Fragment Remasking.** To explore the local chemical neighborhood of x_{init}, we apply a mutation operator termed *Fragment Remasking*. Unlike token-level masking, this operator acts at the semantic level of chemical substructures. A fragment is selected according to a decomposition rule $\mathcal{R}_{\text{remask}}$ and replaced by a sequence of mask tokens $[M]$. The number of mask tokens m is sampled from a predefined distribution p_{len}, such as the empirical fragment-length distribution observed in the training data. This mechanism allows flexible control over the size and complexity of the regenerated substructure. Conceptually, this operation corresponds to Gibbs sampling, where a fragment f_k is resampled from the conditional distribution $p(f_k \mid f_{\setminus k})$, with $f_{\setminus k}$ denoting the unmasked molecular context.

3. **Reconstruction with Molecular Fragment Context.** The masked region is reconstructed

using the discrete diffusion model conditioned on the remaining molecular fragments. Given a partially masked molecule, the diffusion model iteratively denoises the masked positions while attending to the unmasked fragment-level context through self-attention. This conditional reconstruction naturally enforces chemical compatibility between the regenerated fragment and the existing molecular scaffold, as the denoising distribution is explicitly conditioned on the surrounding fragments.

Formally, at each diffusion step t, the model predicts the categorical distribution for selected masked token x^l as

$$x^l_{\theta,i}(z_t, t) = p_\theta(x^l = i \mid z_t), \qquad (17)$$

where z_t denotes the noisy sequence retaining the unmasked fragment context. By sampling from this conditional distribution across diffusion steps, the model generates fragments that are coherent with the molecular structure and aligned with the learned chemical priors. This context-aware reconstruction serves as a structured mutation operator, enabling localized yet chemically valid exploration of the molecular space.

4. **Vocabulary Evolution.** The newly generated molecule x_{new} is evaluated by the task-specific scoring oracle. It is then decomposed into fragments, which are scored using $S(\cdot)$ and merged into the vocabulary. The vocabulary $\mathcal{V}$ is subsequently updated by retaining the top-V fragments from the union of the existing and newly generated candidates. This feedback loop enables continual expansion and refinement of the fragment pool, progressively steering the search toward regions of chemical space associated with higher target property values.

Algorithm 1 Evolutionary Fragment Optimization (EFO)

Input: Dataset of molecules $\mathcal{D}$; vocabulary size V; fragment decomposition rule $\mathcal{R}_{\text{vocab}}$; fragment remasking rule $\mathcal{R}_{\text{remask}}$; number of generations G
Decompose $\mathcal{D}$ into a fragment multiset $\mathcal{F}$ using $\mathcal{R}_{\text{vocab}}$
Initialize fragment vocabulary $\mathcal{V}$ with the top-V fragments from $\mathcal{F}$ ranked by Eq. (16)
Estimate fragment-length distribution p_{len} from $\mathcal{D}$ based on $\mathcal{R}_{\text{remask}}$
Initialize generated molecule set $\mathcal{M} \leftarrow \emptyset$
while $|\mathcal{M}| < G$ **do**
  Sample two fragments from $\mathcal{V}$ and attach them to form an initial molecule x_{init}
  Sample mask length $m \sim p_{\text{len}}$
  Select a fragment in x_{init} according to $\mathcal{R}_{\text{remask}}$
  Replace the selected fragment with m mask tokens to obtain a partially masked molecule x_{mask}
  Reconstruct the masked region via conditional discrete diffusion to obtain x_{new}
  Update $\mathcal{M} \leftarrow \mathcal{M} \cup \{x_{\text{new}}\}$
  Decompose x_{new} into fragments $\{f_1, f_2, \dots\}$ using $\mathcal{R}_{\text{vocab}}$
  Update $\mathcal{V}$ by retaining the top-V fragments from $\mathcal{V} \cup \{f_1, f_2, \dots\}$
end while
Output: Generated molecule set $\mathcal{M}$

G Details of Dual-Conditioned Generation

In this section, we provide detailed experimental settings and quantitative results for the dual-conditioned generation task discussed in Sec. 5.3. As noted in the main text, realistic drug discovery requires candidate molecules to simultaneously achieve strong target binding affinity and acceptable safety profiles. Optimizing binding affinity alone often leads to toxic or developability-limited compounds, while focusing solely on ADMET properties may result in insufficient target engagement. This experiment is designed to evaluate whether CAGenMol can effectively reconcile these heterogeneous and potentially competing objectives within a unified generation framework.

Target protein. We consider the protein structure with PDB ID **3O96**, corresponding to the N-terminal pleckstrin homology (PH) domain of human **AKT1**, a key serine/threonine kinase in the PI3K–AKT signaling pathway. AKT1 is ubiquitously expressed across multiple human tissues and plays a central role in regulating cell survival, metabolism, and proliferation. Dysregulation of AKT1 signaling is closely associated with cancer and metabolic disorders, making it an important therapeutic target. Ligands targeting the AKT1 PH domain must be able to reach intracellular AKT1 while maintaining sufficient selectivity. Moreover, due to the essential physiological functions of AKT1 in normal organs, candidate compounds are required to exhibit low systemic toxicity and acceptable safety profiles, including the absence of mutagenic potential as indicated by a negative Ames test.

Experimental setup. We condition CAGenMol on both the fixed receptor structure of 3O96 and

an intrinsic toxicity constraint requiring generated molecules to be **Ames-negative**. Training follows the same two-stage paradigm used throughout this work. Specifically, we first perform supervised conditional training, after which Step-PPO is applied to optimize a composite reward that jointly accounts for docking affinity (evaluated by AutoDock Vina) and MiniMol-predicted Ames toxicity probability. During inference, we generate **100 molecules** for the target pocket and evaluate each candidate under both structural and ADMET-related metrics.

To better isolate the effect of toxicity conditioning, we additionally report an ablation variant (**w/o Ames Condition**) in which CAGenMol is trained and optimized solely for structure-based binding without enforcing the Ames constraint.

Evaluation metrics. Generated molecules are evaluated using the following criteria: (1) **Vina Dock**: Binding affinity to the AKT1 PH domain; (2) **QED**: Quantitative estimate of drug-likeness; (3) **SA**: Synthetic accessibility score; (4) **Ames**: Predicted mutagenicity probability, where lower values indicate safer compounds. This evaluation protocol reflects a realistic screening scenario in which candidates must simultaneously satisfy potency, developability, and safety requirements.

Results. Quantitative results are summarized in Table 5. Compared with existing structure-conditioned baselines, CAGenMol achieves the lowest predicted Ames toxicity while maintaining competitive docking performance and substantially higher molecular quality in terms of QED and SA. Notably, methods that achieve very strong docking scores often suffer from significantly elevated Ames toxicity, highlighting the limitation of affinity-only optimization.

The comparison between CAGenMol and its ablation variant further demonstrates the effect of explicit toxicity conditioning. While the **w/o Ames Condition** model achieves slightly stronger docking scores, it exhibits nearly twice the Ames toxicity compared to the full model. In contrast, CAGenMol effectively reduces mutagenicity risk with only a marginal trade-off in binding affinity, resulting in a more balanced and practically viable candidate set.

Overall, these results confirm that CAGenMol, empowered by Step-PPO, can successfully coordinate structure-based optimization and safety constraints, avoiding the collapse to single-objective solutions and enabling realistic dual-conditioned molecular generation.

H Evaluation on Additional SBDD Benchmark

Benchmark. We further evaluate CAGenMol on the standardized structure-based drug design benchmark proposed in (Zheng et al., 2024), which uses seven representative target proteins (PDBIDs: 1IEP, 3EML, 3NY8, 4RLU, 4UNN, 5MO4, 7L11) and evaluates top-10 docking performance across diverse algorithmic families.

Protocol. To test generalization across targets, we initialize CAGenMol using the model already fine-tuned on CrossDocked2020 (Sec. 5.1), and then perform additional Step-PPO optimization under each target pocket using the benchmark docking oracle. Following the benchmark setting, we report the average Top-10 docking score for each target and the overall average. Since this benchmark setting naturally supports iterative black-box optimization, we using 3-round EFO at inference time.

Results. Table 3 shows that CAGenMol is competitive with strong optimization-based baselines while maintaining a learned generative prior. We also observe that enabling EFO consistently improves the docking performance by further exploiting high-reward fragment motifs.

I Reward Analysis

To empirically validate the optimization stability and convergence speed of our Step-PPO algorithm, we visualize the reward trajectories during the alignment phase for the three property-conditioned generation tasks described in Section 5.2.

Figure 4-6 illustrates the average reward curves throughout the Step-PPO training process. We observe consistent convergence patterns across all three distinct ADMET settings:

- **Fast Convergence:** In all scenarios, the model effectively learns to navigate the chemical space towards high-reward regions within the first hundred steps. This rapid adaptation demonstrates that the supervised initialization provides a strong chemical prior, allowing Step-PPO to focus immediately on constraint satisfaction rather than relearning basic chemical validity.

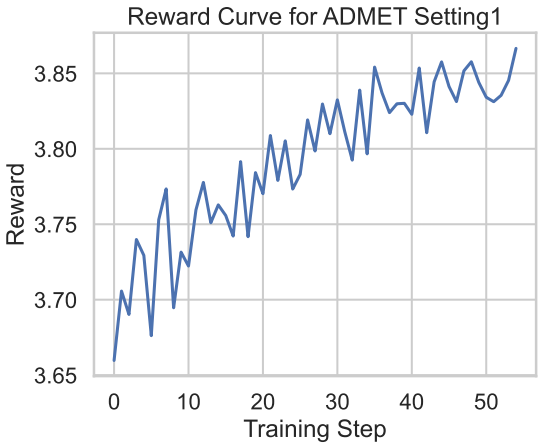

Figure 4: Setting 1 CNS Drugs

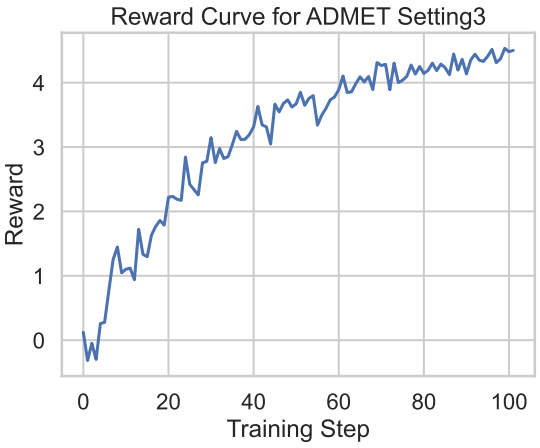

Figure 6: Setting 3 Peripheral Drugs

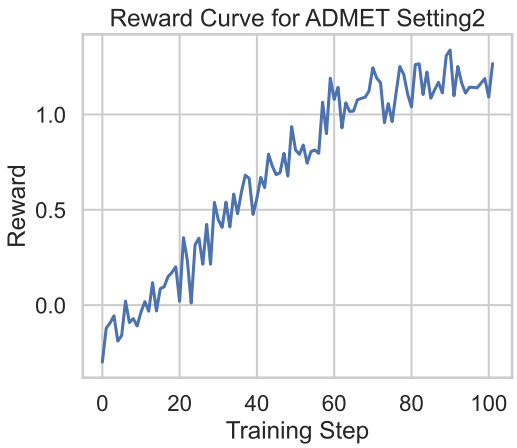

Figure 5: Setting 2 Hepatic Drugs

- **Stability:** Unlike standard RL fine-tuning which often suffers from high variance or collapse, our step-wise formulation maintains a steady ascending trajectory. The relatively narrow variance (if applicable in your plot) suggests that the token-level policy updates are robust and do not degrade the overall structural integrity of the molecules.

- **Task Difficulty:** We note that Setting 2 (Hepatically metabolized drugs) shows a slightly slower convergence rate compared to Setting 1 and 3. This aligns with the complexity of the constraints, as satisfying specific Lipophilicity ranges combined with enzyme substrate specificity represents a more constrained optimization landscape.

These training dynamics confirm that CAGen-Mol can efficiently align the generative distribution with complex, multi-objective property constraints without requiring extensive hyperparameter tuning or suffering from mode collapse.

J Ablation Study

In this section, we investigate the contribution of the specific architectural designs in the Unified Constraint Adaptor (UCA) and the impact of the reinforcement learning stage.

Note that the effectiveness of the Evolutionary Fragment Optimization (EFO) has already been demonstrated in Section 5.2 and Appendix H, where enabling EFO consistently improved performance. Therefore, we exclude EFO from this analysis.

We compare five variants to dissect the framework:

- **SFT (Base):** A baseline supervised model where the UCA uses mean pooling instead of linear attention, and receives only ESM-2 embeddings without explicit physicochemical features.

- **SFT (w/o Attn.):** UCA uses mean pooling, but includes both ESM-2 and physicochemical features.

- **SFT (w/o Phys.):** UCA uses linear attention, but relies solely on ESM-2 embeddings (no physicochemical stream).

- **SFT (Full UCA):** The complete UCA architecture (Attention + Phys. Features) trained only with Supervised Fine-Tuning (no RL).

- **Full Model (Step-PPO):** The complete CAGenMol framework fine-tuned with Step-wise PPO.

The results on the CrossDocked2020 test set are summarized in Table 2.

Architecture Analysis (Impact of UCA Design). Comparing the SFT variants reveals the importance of our UCA design. First, **SFT (w/o Phys.)** generally underperforms **SFT (Full UCA)**, indicating that while protein language models provide rich semantics, the explicit physicochemical features (charge, hydropathy, etc.) provide critical guidance for precise surface matching. Second, replacing the linear attention with mean pooling (**SFT w/o Attn.**) leads to a performance drop. This suggests that the attention mechanism successfully learns to weigh critical residues in the pocket, whereas mean pooling dilutes the signal from key binding sites. The **SFT (Base)** variant, lacking both designs, yields the lowest performance among the supervised models, validating the synergy of our dual-stream encoder and attention pooling.

Impact of Reinforcement Learning. A significant performance gap is observed between **SFT (Full UCA)** and the **Full Model (Step-PPO)**. While the supervised model with the full UCA architecture achieves reasonable validity and docking scores, it fails to reach the high-affinity regime. The introduction of Step-PPO drastically improves the Vina Dock score and Success Rate without collapsing diversity. This confirms that while the UCA provides a necessary condition-aware representation, the Step-wise PPO algorithm is indispensable for aligning the generation process with the stringent binding affinity objectives.

Extended Ablation: Role of Initialization, RL, and EFO. To further address the contribution of each component raised by the reviewer, we conduct additional controlled experiments on pocket 3o96_A due to computational constraints. Specifically, we investigate whether the performance gains mainly stem from reinforcement learning (Step-PPO), and whether similar improvements can be achieved by applying Step-PPO or EFO on weaker initializations.

Table 6 shows the results. We compare: (i) the full model initialized from SFT (Full UCA), (ii) a weaker initialization SFT (Base) followed by Step-PPO, (iii) Step-PPO without any SFT initialization, and (iv) EFO-only variants.

These results provide several key insights. First, while Step-PPO significantly improves performance over supervised models, its effectiveness depends strongly on the initialization. Removing SFT leads to both degraded docking performance and substantially slower convergence, indicating that SFT provides a crucial chemical prior and stabilizes RL optimization.

Second, applying Step-PPO on a weaker backbone (SFT Base) narrows but does not close the gap with the full model, suggesting that architectural improvements and RL contribute complementary gains. This also clarifies that our method is not equivalent to simply fine-tuning GenMol with RL, as the conditioning-aware UCA architecture remains essential.

Third, EFO alone yields limited improvements even with increased iterations, due to the extremely large combinatorial search space. This confirms that EFO acts as a refinement module rather than a standalone optimizer, and cannot replace the generative policy learned via RL.

Overall, these findings demonstrate that (1) SFT is critical for stable and efficient RL optimization, (2) Step-PPO is the primary driver for achieving high-affinity generation, and (3) EFO provides additional but limited gains through local refinement.

K Inference Efficiency and Runtime Analysis

In addition to generation performance, inference efficiency is an important practical factor for drug design. Frag2Seq (Fu et al., 2024) reports the wall-clock time required to generate 100 molecules per pocket for a range of structure-conditioned molecular generation methods. Following the same evaluation protocol, we compare CAGenMol with these representative baselines.

We run on a single NVIDIA A800 GPU with a batch size of 100. We report the inference time for the base model as well as the variant augmented with EFO using 3 refinement rounds.

Table K summarizes the runtime comparison. CAGenMol achieves orders-of-magnitude faster inference than diffusion-based and graph-based baselines. Even with EFO enabled, CAGenMol remains significantly faster than all compared methods, demonstrating its suitability for large-scale and iterative molecular generation scenarios.

Method	Vina Dock (↓)	QED (↑)	SA (↑)
SFT(Full)+Step-PPO	**-10.13**	0.83	0.88
SFT(Base)+Step-PPO	-9.97	**0.84**	0.88
Step-PPO only	-9.49	0.84	**0.89**
EFO only (3 iters)	-7.66	0.70	0.75
EFO only (5 iters)	-8.02	0.69	0.77

Table 6: Ablation of initialization, RL fine-tuning, and EFO on pocket 3o96_A.

Method	Time (s)
3D-SBDD (Luo et al., 2021)	15986.4
Pocket2Mol (Peng et al., 2022)	2827.3
GraphBP (Liu et al., 2022)	1162.8
TargetDiff (Guan et al., 2023)	3428.0
DecompDiff (Guan et al., 2024)	6189.0
DiffSBDD (Schneuing et al., 2024)	629.9
FLAG (Zhang et al., 2023)	1289.1
DrugGPS (Zhang and Liu, 2023)	1007.8
Lingo3DMol (Feng et al., 2024)	1481.9
Frag2Seq (Fu et al., 2024)	48.8
CAGenMol	**3.5**
CAGenMol + EFO (3 rounds)	**29.9**