TRAINING-FREE STEIN DIFFUSION GUIDANCE: POSTERIOR CORRECTION FOR SAMPLING BEYOND HIGH-DENSITY REGIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Training-free diffusion guidance provides a flexible way to leverage off-the-shelf classifiers without additional training. Yet, current approaches hinge on posterior approximations via Tweedie's formula, which often yield unreliable guidance, particularly in low-density regions. Stochastic optimal control (SOC), in contrast, provides principled posterior simulation but is prohibitively expensive for fast sampling. In this work, we reconcile the strengths of these paradigms by introducing Stein Diffusion Guidance (SDG), a novel training-free framework grounded in a surrogate SOC objective. We establish a theoretical bound on the value function, demonstrating the necessity of correcting approximate posteriors to faithfully reflect true diffusion dynamics. Leveraging Stein variational inference, SDG identifies the steepest descent direction that minimizes the Kullback-Leibler divergence between approximate and true posteriors. By incorporating a principled Stein correction mechanism and a novel running cost functional, SDG enables effective guidance in low-density regions. Experiments on molecular low-density sampling tasks suggest that SDG consistently surpasses standard training-free guidance methods, highlighting its potential for broader diffusion-based sampling beyond high-density regions.

1 INTRODUCTION

In many scientific domains, key discoveries often depend on identifying rare samples buried within large data distributions. For instance, while billions of molecules exist in chemistry (Polishchuk et al., 2013), only a minute fraction possesses properties relevant to drug discovery. We posit that such high-value samples frequently reside in low-density regions, making their identification both difficult and error-prone. This challenge has fueled growing interest in methods that accelerate the search for rare, property-rich samples. Generative methods, particularly diffusion models (Ho et al., 2020; Song et al., 2021), have demonstrated strong performance in modeling complex, high-dimensional distributions. However, when trained on unlabeled data, diffusion models predominantly sample from high-density data regions, thereby overlooking the low-density areas where high-value samples are likely to exist. This limitation hinders their effectiveness in tasks that require discovery beyond high-density regions. Numerous studies have been proposed to address this challenge. A particu-

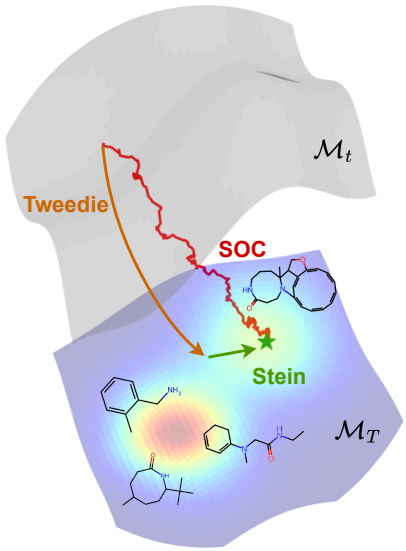

Figure 1: SDG provides a computationally efficient alternative to SOC-based diffusion guidance for molecular sampling in low-density regions.

lar class of methods leverages an auxiliary classifier, (Dhariwal & Nichol, 2021; Sehwag et al., 2022; Lee et al., 2023), to guide pretrained diffusion models toward regions of interest. However, classifier-based diffusion guidance introduces additional complexity by requiring classifiers trained on multiple noise levels. Moreover, recent studies suggest that noisy classifier gradients can misguide samples, causing them to fall off generative manifolds (Chung et al., 2023; Guo et al., 2024). This issue can be even severe in low-data regimes, where diffusion models are already less accurate and gradients tend to be less reliable (Sehwag et al., 2022).

Stochastic optimal control (SOC) (Nüsken & Richter, 2021; Domingo-Enrich et al., 2024) has recently been explored to fine-tune diffusion models for a variety of downstream tasks (Uehara et al., 2024; Wang et al., 2025; Domingo-Enrich et al., 2025). These approaches steer the diffusion process towards desired targets by incorporating an auxiliary controller into the stochastic differential equation (SDE) that governs the generative reverse diffusion dynamics. Uehara et al. (2024) further relate SOC to classifier-based guidance, where the reward functions are classifiers trained on clean data, which are readily available across many domains. This enables SOC-based diffusion guidance to leverage off-the-shelf classifiers directly. However, computing the optimal control value requires backpropagating reward signals through entire neural-SDE sampling trajectories (Tzen & Raginsky, 2019; Uehara et al., 2024), which presents a significant drawback restricting the scalability and practicality of SOC-based sampling methods. To circumvent this problem, recent works have proposed approximating the diffusion posterior through Tweedie's formula (Robbins, 1992). This avenue has been primarily explored in the contexts of general inverse problems (Chung et al., 2023; MOUFAD et al., 2025), and image diffusion finetuning applications (Yu et al., 2023; Ma et al., 2024; Rout et al., 2025; Janati et al., 2025; Dinh et al., 2025). These methods are often referred to as *training-free diffusion guidance*, as they leverage off-the-shelf classifiers without requiring additional training across noise levels. However, the posterior approximation via Tweedie's formula is biased and inherently suboptimal, which frequently leads to unreliable guidance, particularly in low-density data regions.

Here, we summarize our contributions: (i) we propose a low-density diffusion guidance framework formulated under stochastic optimal control, which introduces a novel cost-to-go function; (ii) we theoretically prove that approximating the diffusion posterior via Tweedie's formula is inferior and requires further correction steps; (iii) we introduce a Stein correction mechanism for surrogate SOC-based diffusion sampling, which leverages Stein variational inference tools (Liu & Wang, 2016; Liu, 2017) to iteratively minimize the Kullback-Leibler (KL) divergence between approximate and true posteriors. Our experimental results suggest that the proposed Stein correction is critical for enhancing training-free diffusion guidance in molecular sampling problems, particularly in low-density regions, leading to the discovery of molecules with higher binding affinities to target proteins.

2 PRELIMINARIES

2.1 DIFFUSION MODELS

Song et al. (2021) introduce a continuous-time, continuous-state diffusion framework, $\forall t \in [0, T]$, $\mathbf{x} \in \mathbb{R}^d$, utilizing a pair of forward and backward SDEs. The forward process diffuses data samples, $\mathbf{x}_T \sim p_T$, towards an easy-to-sample prior, $\mathbf{x}_0 \sim p_0$. Its dynamic is the solution to an Itô SDE: $d\mathbf{x}_t = \mathbf{b}(\mathbf{x}_t, t)dt + \sigma(t)d\mathbf{w}$, where $dt < 0$ is an infinitesimal timestep, $\mathbf{w}$ is the Weiner process, and $\mathbf{b}(\mathbf{x}_t, t), \sigma(t)$ denote the drift and diffusion coefficient, respectively. A backward process gradually denoises samples from the prior p_0 and back to the data distribution p_T; the reverse dynamic corresponds to another Itô SDE of the form:

$$\mathbb{P}: \quad d\mathbf{x}_t = \left(-\mathbf{b}(\mathbf{x}_t, t) + \sigma(t)^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)\right) dt + \sigma(t) d\mathbf{w} \tag{1}$$

where the marginal data score, $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$, can be estimated by the score-matching technique (Hyvärinen & Dayan, 2005) via a time-dependent score-based network, $\mathbf{s}_\theta(\mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t)$; for simplicity, we omit the explicit t-dependence in the score model notation. Unlike Song et al. (2021), we employ a positive infinitesimal reverse timestep to facilitate our theoretical development of novel diffusion guidance in subsequent sections. In sampling, diffusion models can incorporate a classifier $r(\cdot)$ to guide samples toward regions with desired properties, a technique known as *classifier-based diffusion guidance* (Dhariwal & Nichol, 2021). The conditional score can be factorized into $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|y) \propto \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} r(y|\mathbf{x}_t)$, which requires training the classifier $r(\cdot)$ on noisy data of multiple noise levels, $\mathbf{x}_t \sim p_t(\mathbf{x}_t|\mathbf{x}_T), \forall t \in [0, T]$.

2.2 STEIN VARIATIONAL GRADIENT DESCENT

Variational inference (VI) approximates a complex target distribution p using a simpler tractable distribution $q \in \mathcal{Q}$, by minimizing the KL divergence, $\arg\min_q D_{KL}(q\|p)$. Stein variational gradient descent (SVGD) (Liu & Wang, 2016) provide a nonparametric approach, representing q as a set of interacting particles that are deterministically evolved along a direction ϕ to most efficiently decrease the KL divergence. Liu & Wang (2016) solve the steepest descent problem in the unit ball of reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. The following lemma summarizes their core result:

Lemma 2.1 *$-\nabla_{\mathbf{x}} D_{KL}(q\|p)$ steepest direction maximizes Stein discrepancy in the RKHS unit ball.*

$$\phi^* = \arg\max_{\phi\in\mathcal{H}} \mathbb{E}_{\mathbf{x}\sim q}\left[\textit{trace}\left(\mathcal{A}_p\phi(\mathbf{x})\right)\right] \quad \textit{st} \quad \|\phi\|_{\mathcal{H}} \le 1$$

$$\textit{with} \quad \mathcal{A}_p\phi(\mathbf{x}) = \nabla_{\mathbf{x}} \log p(\mathbf{x})\phi(\mathbf{x}) + \nabla_{\mathbf{x}}\phi(\mathbf{x})$$

where $\mathcal{A}_p$ is the Stein operator, ϕ is the Stein class of the target density p. Liu et al. (2016) solve the problem 2.1 as kernelized Stein discrepancy (KSD) that yields $\phi^*(\mathbf{x}^i) = \mathbb{E}_{\mathbf{x}^j\sim q(\mathbf{x})}\left[\nabla_{\mathbf{x}^j}\log p(\mathbf{x}^j)k(\mathbf{x}^i,\mathbf{x}^j) + \nabla_{\mathbf{x}^j}k(\mathbf{x}^i,\mathbf{x}^j)\right]$, where $k(\mathbf{x}^i,\mathbf{x}^j)$ is Radial Basic Function (RBF) kernel, $k(\mathbf{x}^i,\mathbf{x}^j) = \exp\left(-\frac{1}{h}\|\mathbf{x}^i - \mathbf{x}^j\|_2^2\right)$. SVGD can approximate the intractable target density p by gradually transporting a set of N initial particles $\{\mathbf{x}_0^i\}_{i=0}^N \sim q_0$ along the direction ϕ^*, which relies on the computable score $\nabla_{\mathbf{x}}\log p(\mathbf{x})$. The first term of ϕ^* represents a kernel-weighted gradient ascent direction that pushes the particles toward the high-density regions. The second term denotes the repulsive force that prevents the particles from collapsing into the local modes of $p(\mathbf{x})$.

2.3 STOCHASTIC OPTIMAL CONTROL

Stochastic optimal control (Nüsken & Richter, 2021) seeks an optimal controller that steers the behavior of a given stochastic system to minimize a pre-specified cost function. For the stochastic dynamical diffusion system in Equation 1, we formalize an affine-control problem as follows:

$$\inf_{\mathbf{u}\in\mathcal{U}} \mathbb{E}\left[\int_t^T \left(\frac{1}{2}\|\mathbf{u}(\mathbf{x}_t^{\mathbf{u}},t)\|^2 + f(\mathbf{x}_t^{\mathbf{u}},t)\right)dt + g(\mathbf{x}_T^{\mathbf{u}})\right]$$

$$s.t. \quad \mathbb{P}^{\mathbf{u}}: \quad d\mathbf{x}_t^{\mathbf{u}} = \left(-\mathbf{b}(\mathbf{x}_t^{\mathbf{u}},t) + \sigma(t)^2\nabla_{\mathbf{x}_t^{\mathbf{u}}}\log p_t(\mathbf{x}_t^{\mathbf{u}}) + \sigma(t)\mathbf{u}(\mathbf{x}_t^{\mathbf{u}},t)\right)dt + \sigma(t)d\mathbf{w} \qquad (2)$$

where the feedback control $\mathbf{u}: \mathbb{R}^d \times [0,T] \mapsto \mathbb{R}^d$ drives the system dynamics, $f: \mathbb{R}^d \times [t,T] \mapsto [0,\infty)$ specifies the state cost, $g: \mathbb{R}^d \mapsto [0,\infty)$ is the terminal cost, and $\mathbb{P}^u$ denotes the controlled probability path measure induced from $\mathbb{P}$. The control objective minimizes the cost functional $J(\mathbf{u},\mathbf{x},t) = \mathbb{E}_{\mathbb{P}^{\mathbf{u}}}\left[\int_t^T \left(\frac{1}{2}\|\mathbf{u}(\mathbf{x}_s^{\mathbf{u}},s)\|^2 + f(\mathbf{x}_s^{\mathbf{u}},s)\right)ds + g(\mathbf{x}_T^{\mathbf{u}})|\mathbf{x}_t = \mathbf{x}\right]$, whose minimum defines the value function or optimal cost-to-go (Fleming & Soner, 2006), $V(\mathbf{x},t) = \inf_{\mathbf{u}\in\mathcal{U}} J(\mathbf{u},\mathbf{x},t)$. Moreover, verification theorem (Fleming & Soner, 2006; Pham, 2009) relates the optimal control and value function via $\mathbf{u}^* = -\sigma\nabla_{\mathbf{x}}V$. The affine stochastic control problem with a quadratic cost is closely connected to an iterative diffusion optimization using a relative entropy loss (Powell, 2021; Kappen et al., 2012; Hartmann & Schütte, 2012), which involves simulating multiple controlled trajectories, computing their cumulative costs, and backpropagating through the trajectories to update a parameterized controller; a more extensive treatment of SOC problem can be found in (Nüsken & Richter, 2021; Domingo-Enrich et al., 2024). Here, we summarize their essential results as follows:

Lemma 2.2 *Optimal controller and value function for the control problem in Equation 2.*

$$\mathbf{u}^*(\mathbf{x},t) = -\sigma(t)\nabla_{\mathbf{x}}V(\mathbf{x},t)$$

$$V(\mathbf{x},t) = -\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\int_t^T f(\mathbf{x}_s,s)\,ds - g(\mathbf{x}_T)\right)|\mathbf{x}_t = \mathbf{x}\right]$$

As observed, obtaining the value function requires integrating diffusion trajectories from $\mathbf{x}_t = \mathbf{x}$ under $\mathbb{P}$, and computing the gradients $\nabla_{\mathbf{x}}V(\mathbf{x},t)$ within these simulation trajectories to recover the optimal controller. Both operations are computationally expensive and substantially slow for practical applications. Moreover, the state cost is often omitted, i.e, $f(\mathbf{x}_s,s) = 0\ \forall s$. Under this setting, Uehara et al. (2024) establish a connection with classifier-based diffusion guidance, $\mathbf{u}^* = \sigma(t)\nabla_{\mathbf{x}}\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(-g(\mathbf{x}_T)\right)|\mathbf{x}_t = \mathbf{x}\right] \propto \nabla_{\mathbf{x}}\mathbb{E}_{\mathbb{P}}\left[r(\mathbf{x}_T)|\mathbf{x}_t = \mathbf{x}\right]$, wherein $r = -g$, due to the minimization problem, corresponds to an off-the-shelf classifier or a differential reward model.

3 STEIN DIFFUSION GUIDANCE

We introduce a novel training-free diffusion guidance framework derived from a surrogate stochastic optimal control (SOC) formulation. Section 3.1 presents a new SOC cost functional that enables diffusion models to explore low-density regions. Section 3.2 establishes a variational bound on the SOC value function, showing that existing training-free guidance methods require posterior correction. Section 3.3 proposes a back-and-forth Stein correction, a low-cost alternative to SOC that regularizes posterior samples for effective low-density exploration. Detailed proofs are deferred to Appendix B.

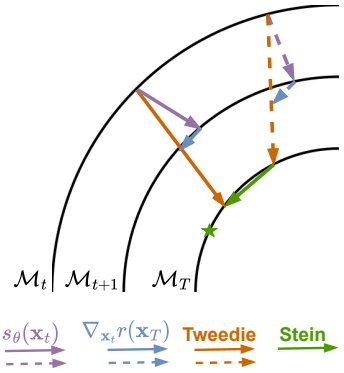

Figure 2: Back-and-forth Stein correction: Particles are mapped backward to $\mathcal{M}_T$ to obtain posterior samples, which are corrected via Stein correction, and then mapped forward to $\mathcal{M}_t$ for reward-based guidance. Dashed arrows indicate the standard training-free method, while solid arrows denote SDG.

3.1 LOW DENSITY DIFFUSION SAMPLING AS STOCHASTIC OPTIMAL CONTROL

We consider the controlled reverse SDE (Equation 2) and its associated probability path measure $\mathbb{P}^{\mathbf{u}}$. We introduce a novel cost functional $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ that progressively anneals the marginal density $p_t(\mathbf{x}_t)$ of the uncontrolled SDE (Equation 1) under $\mathbb{P}$ to low-density regions:

$$\widetilde{J}(\mathbf{u}, \mathbf{x}, t) = \mathbb{E}_{\mathbb{P}^{\mathbf{u}}}\left[\int_t^T \left(\frac{1}{2}\|\mathbf{u}(\mathbf{x}_s^{\mathbf{u}}, s)\|^2 + \alpha(s)\log p_s(\mathbf{x}_s^{\mathbf{u}})\,\delta(s-t)\right) ds - \beta(t) r(\mathbf{x}_T^{\mathbf{u}}) | \mathbf{x}_t = \mathbf{x}\right] \tag{3}$$

where $\alpha(t)$ and $\beta(t)$ denote the schedules controlling low-density annealing and guidance strength, respectively. We formulate low-density diffusion guidance as a stochastic optimal control problem.

Proposition 3.1 *Consider the SOC problem in Equation 2 with the novel functional cost $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ defined in Equation 3. By Lemma 2.2, the marginal density $p_t^{\mathbf{u}}(\mathbf{x}_t)$, the value function $V(\mathbf{x}, t)$, and the optimal control $\mathbf{u}^*(\mathbf{x}, t)$ of the controlled-reverse SDE under $\mathbb{P}^{\mathbf{u}}$ are given as*

$$p_t^{\mathbf{u}}(\mathbf{x}_t) = p_t^{1-\alpha(t)}(\mathbf{x}_t)\exp(\beta(t) r(\mathbf{x}_T))$$
$$\mathbf{u}^*(\mathbf{x}, t) = \sigma(t)\nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \quad \textit{and} \quad V(\mathbf{x}, t) = -\log \mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right]$$

The induced marginal density is the product of the annealed density term $p_t^{1-\alpha(t)}$ and the guidance term $\exp(\beta(t) r(\mathbf{x}_T))$. The latter is an un-normalized energy density with the energy function $-\beta(t) r(\mathbf{x}_T)$. In Section 2.3, SOC constitutes a computationally expensive simulation problem. In particular, obtaining the optimal control $\mathbf{u}^*(\mathbf{x}, t)$ requires backpropagating through sampling trajectories (Uehara et al., 2024; Wang et al., 2025). To alleviate the computational burden, one can directly approximate posterior samples via Tweedie's formula $\mathbf{x}_T \approx ((\mathbf{x}_t + \gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t, t))/\eta(t)$, assuming a forward kernel $p_{t|T}(\mathbf{x}_t|\mathbf{x}_T) = \mathcal{N}(\eta(t)\mathbf{x}_T, \gamma^2(t) I)$. This one-step approximation has been explored in many prior works. However, using Tweedie-based proposal distribution is inherently suboptimal. As illustrated in Figure 1, the Tweedie-based approximation significantly deviates from the endpoint sample given by the SOC simulation. In the following sections, we analyze the sub-optimality of the Tweedie-based posterior and propose a novel Stein correction mechanism.

3.2 VALUE FUNCTION VARIATIONAL BOUND

We consider the target posterior $p_{T|t}(\mathbf{x}_T \mid \mathbf{x}_t)$, defined as the terminal distribution of $\mathbf{x}_t$ under the uncontrolled process $\mathbb{P}$ from Equation 1. By Proposition 3.1, the low-density reward-guided value function can be written as $V(\mathbf{x}, t) = -\log \mathbb{E}_{\mathbf{x}_T \sim p_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right]$, where $p_t^{\mathbf{u}}(\mathbf{x}) = p_t^{1-\alpha(t)}(\mathbf{x})\exp(\beta(t) r(\mathbf{x}_T))$. To ensure computational tractability, we introduce a surrogate objective $\bar{V}(\mathbf{x}, t, q)$ with $q \in \mathcal{Q}$, which serves as an upper bound on the value function $V(\mathbf{x}, t)$.

Proposition 3.2 *Let $q \in \mathcal{Q}$ be any traceable family of proposal distributions. Then, the value function in Proposition 3.1 admits the following upper bound:*

$$\begin{aligned} V(\mathbf{x},t) &\leq \bar{V}(\mathbf{x},t,q) \\ &= -\mathbb{E}_{\mathbf{x}_T\sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[\log\left(\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})}\right)|\mathbf{x}_t=\mathbf{x}\right] + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t)\|p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t=\mathbf{x}} \\ &= \alpha(t)\log p_t(\mathbf{x}) - \beta(t)\mathbb{E}_{\mathbf{x}_T\sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[r(\mathbf{x}_T)|\mathbf{x}_t=\mathbf{x}\right] \\ &\qquad + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t)\|p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t=\mathbf{x}} \end{aligned}$$

The first term on the RHS drives samples toward low-density regions, while the second term guides them toward regions with desired properties. When posterior samples $\mathbf{x}_T$ are estimated via Tweedie's formula, these two terms together reproduce the training-free diffusion guidance explored in prior works (Chung et al., 2023; MOUFAD et al., 2025; Yu et al., 2023; Ma et al., 2024; Shen et al., 2024; Rout et al., 2025; Janati et al., 2025; Dinh et al., 2025). The last term serves as a KL regularization that minimizes the divergence between the proposal posterior $q(\mathbf{x}_T|\mathbf{x}_t)$ and the true posterior $p(\mathbf{x}_T|\mathbf{x}_t)$. This reveals that optimizing only the first two terms of $\bar{V}(\mathbf{x},t,q)$ under the Tweedie-based approximation is suboptimal, since the resulting controller neglects the KL term and thus fails to remain close to the true posterior. To overcome this limitation, we introduce a Stein correction mechanism that refines the approximate diffusion posterior before sampling for guidance.

3.3 Stein meets Tweedie for surrogate stochastic optimal control

Given the variational upper bound of the value function, we derive the optimal controller that minimizes this bound, which we term the *surrogate* stochastic optimal control problem. By Lemma 2.2, the optimal control $\bar{\mathbf{u}}^*(\mathbf{x}_t,t)$ for the surrogate value function $\bar{V}(\mathbf{x},t,q)$ can be obtained as follows:

$$\begin{aligned} &\frac{\bar{\mathbf{u}}^*(\mathbf{x}_t,t)}{\sigma(t)} = -\nabla_{\mathbf{x}_t}\bar{V}(\mathbf{x}_t,t,q) \\ &= \underbrace{-\alpha(t)\mathbf{s}_\theta(\mathbf{x}_t) + \beta(t)\nabla_{\mathbf{x}_t}\mathbb{E}_{\mathbf{x}_T\sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[r(\mathbf{x}_T)\right]}_{\mathbf{I}} + \underbrace{-\nabla_{\mathbf{x}_t}D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t)\|p(\mathbf{x}_T|\mathbf{x}_t)\right)}_{\mathbf{II}} \end{aligned} \tag{4}$$

As observed, the first control component (I) recovers the standard training-free diffusion guidance, where the proposal posterior mean is approximated via Tweedie's formula. To address the previously discussed limitations of such methods, we introduce an auxiliary control component (II) that enforces proximity between the proposal and true posteriors, ensuring $q(\mathbf{x}_T|\mathbf{x}_t) \approx p(\mathbf{x}_T|\mathbf{x}_t)$. However, since the true posterior has no closed-form expression, evaluating the KL term directly is infeasible. To address this, we adopt a particle-based optimization strategy using Stein variational inference. Given a set of N particles $\mathcal{D}_t \leftarrow \{\mathbf{x}_t^i\}_{i=0}^N$ on the diffusion manifold $\mathcal{M}_t$, we evolve them along the KL-minimizing direction, which follows from Lemma 2.1:

$$\phi^*(\mathbf{x}_t^i) = \mathbb{E}_{\mathbf{x}_T^j\sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}[\nabla_{\mathbf{x}_t^j}\log p(\mathbf{x}_T^j|\mathbf{x}_t^j)k(\mathbf{x}_T^i,\mathbf{x}_T^j) + \nabla_{\mathbf{x}_t^j}k(\mathbf{x}_T^i,\mathbf{x}_T^j)] \tag{5}$$

We *initialize* the proposal posterior $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ by a new set of particles via Tweedie's formula, i.e, $\mathcal{D}_T \leftarrow \left\{\mathbf{x}_T^i|\mathbf{x}_T^i = \frac{\mathbf{x}_t^i+\gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t^i)}{\eta(t)}, \forall \mathbf{x}_t^i \in \mathcal{D}_t\right\}$. However, directly computing this optimal control direction requires numerous Jacobian-vector products, which are memory-intensive in high-dimensional cases. To solve this computation burden, we propose a back-and-forth Stein correction.

Back-and-forth Stein correction. We first map the particles $\mathcal{D}_t$ *backward*, $\mathcal{M}_t \to \mathcal{M}_T$, to obtain $\mathcal{D}_T$, then apply the Stein correction on $\mathcal{D}_T$, and finally map the corrected particles *forward*, $\mathcal{M}_T \to \mathcal{M}_t$, back to the noisy manifold. We slightly abuse the terms forward and backward (or reverse) here, following the conventions of score-based diffusion models (Ho et al., 2020; Song et al., 2020b). Figure 2 illustrates the key steps of this back-and-forth Stein correction. The steepest descent direction for minimizing $-\nabla_{\mathbf{x}_T}D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t)\|p(\mathbf{x}_T|\mathbf{x}_t)\right)$ on $\mathcal{M}_T$ can thus be computed more efficiently.

$$\phi^*(\mathbf{x}_T^i) = \mathbb{E}_{\mathbf{x}_T^j\sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}[\nabla_{\mathbf{x}_T^j}\log p(\mathbf{x}_T^j|\mathbf{x}_t^j)k(\mathbf{x}_T^i,\mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j}k(\mathbf{x}_T^i,\mathbf{x}_T^j)] \tag{6}$$

As discussed, the true posterior $p(\mathbf{x}_T|\mathbf{x}_t)$ has no closed-form expression; however, its score $\nabla_{\mathbf{x}_T}\log p(\mathbf{x}_T|\mathbf{x}_t)$ can be approximated using score models, as established by the following result.

Lemma 3.3 *For $\mathbf{x}_T \sim p(\mathbf{x}_T|\mathbf{x}_t)$, the posterior score admits the following approximation in terms of the score model $\mathbf{s}_{\boldsymbol{\theta}}(\cdot)$:*

$$\nabla_{\mathbf{x}_T} \log p(\mathbf{x}_T|\mathbf{x}_t) \approx \mathbf{s}_\theta(\mathbf{x}_T) - \eta(t)\mathbf{s}_\theta(\mathbf{x}_t)$$

Here, we assume a forward-diffusion kernel of the form $p_{t|T}(\mathbf{x}_t|\mathbf{x}_T) = \mathcal{N}(\eta(t)\mathbf{x}_T, \gamma^2(t)I)$. Applying the posterior score approximation to Equation 6 yields the evolving direction $\phi^*(\mathbf{x}_T^i)$ of the particles on $\mathcal{M}_T$. We now present the optimal controller for the surrogate stochastic optimal control.

Proposition 3.4 *Consider the low-density reward-based cost functional $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ and its upper bound value function $\bar{V}(\mathbf{x}, t, q)$. Let $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ denote the proposal posterior initialized via Tweedie's formula, and let $q^\epsilon_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ denote the updated posterior obtained after applying the back-and-forth Stein correction with step size $\epsilon(t)$. Then, the optimal control $\bar{\mathbf{u}}^*(\mathbf{x}, t)$ for the surrogate value function $\bar{V}(\mathbf{x}, t, q)$ decomposed as*

$$\frac{\bar{\mathbf{u}}^*(\mathbf{x}_t^i, t)}{\sigma(t)} = \underbrace{-\alpha(t)\mathbf{s}_\theta\left(\mathbf{x}_t^i\right) + \beta(t)\nabla_{\mathbf{x}_t^i}\mathbb{E}_{\mathbf{x}_T^i \sim q^\epsilon_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[r(\mathbf{x}_T^i)\right]}_{\textbf{\textit{Low-density reward-based guidance on } } \mathcal{M}_t}$$

$$\oplus \quad \underbrace{\mathbb{E}_{\mathbf{x}_T^j \sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[\left(\mathbf{s}_\theta(\mathbf{x}_T^j) - \eta(t)\mathbf{s}_\theta(\mathbf{x}_t^j)\right)k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j}k(\mathbf{x}_T^i, \mathbf{x}_T^j)\right]}_{\textbf{\textit{Stein diffusion posterior correction on } } \mathcal{M}_T}$$

Here, $\oplus$ denotes the concatenation operator. The control factor associated with the Stein correction on $\mathcal{M}_T$ refines the initial proposal distribution $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ toward the true diffusion posterior, incurring the updated posterior $q^\epsilon_{T|t}(\mathbf{x}_T|\mathbf{x}_t) \approx p_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$. After mapping the corrected particles forward to the noisy manifold $\mathcal{M}_t$, the second control factor guides the particles toward low-density regions with desired properties. Crucially, our Stein correction ensures robust and accurate guidance even when leveraging off-the-shelf classifiers on posterior samples approximated via Tweedie's formula, thereby improving the method's robustness across diverse tasks. Figure 2 shows that standard training-free guidance often drifts samples outside generative manifolds, whereas Stein-corrected posteriors guarantee reliable guidance that keeps samples within them. We refer to the proposed method as Stein Diffusion Guidance (SDG), which we summarize in Algorithm 1 in the Appendix.

Generalization of Langevin correction. We employ an adaptive step size $\epsilon(t)$ for the Stein correction, with its formulation provided in Appendix C.2. Notably, in the limit $\epsilon(t) \to 0$, the back-and-forth Stein correction recovers the Langevin correction of Song et al. (2020b).

Corollary 3.5 *Let the correction stepsize be set to zero, $\epsilon(t) = 0$ for all t, then the back-and-forth Stein correction reduces to the Langevin correction with stepsize $\gamma^2(t)$ and noise scaled by $\sqrt{2}$:*

$$\mathbf{x}_t \leftarrow \mathbf{x}_t + \gamma^2(t)s_\theta(\mathbf{x}_t) + \gamma(t)\mathbf{z}, \qquad \mathbf{z} \sim \mathcal{N}(0, I)$$

Moreover, the Stein correction incorporates interactions between particles arising from repulsive forces, which play a nontrivial role in enhancing the guidance strength toward desired targets. Similar interaction forces have been used for non-i.i.d. diverse diffusion sampling (Corso et al., 2024).

Table 1: Comparison of training-free diffusion guidance methods on *non-low-density* image guidance tasks. Relevant baseline results are taken from Ye et al. (2024).

Method	Label Guidance		Gaussian Deblur		Super Resolution	
	Accuracy (%) ↑	FID ↓	LPIPS ↓	FID ↓	LPIPS ↓	FID ↓
DPS (Chung et al., 2023)	50.1	172.0	0.390	98.30	0.420	109.0
LGD (Song et al., 2023)	32.2	102	0.270	85.1	0.360	96.7
SDG w/o Stein correction	48.2	89.50	0.326	87.23	0.315	91.99
SDG ($\alpha(t) = 0$, $\epsilon(t) > 0$)	54.0	105.4	0.246	70.00	0.228	68.90

4 EXPERIMENTS

We evaluate SDG on two training-free diffusion guidance settings: (i) standard image diffusion guidance tasks without low-density sampling to assess its generalizability across data domains (Section 4.1); and (ii) reward-guided molecular diffusion sampling in low-density regions (Section 4.2).

4.1 TRAINING-FREE DIFFUSION GUIDANCE ON GENERAL IMAGE TASKS

We adapt the benchmarks from Ye et al. (2024) for three tasks: image label guidance, Gaussian deblurring, and super-resolution. Detailed task descriptions, evaluation metrics, and baselines are provided in Appendix D. Since these tasks do not contain minority class samples, we disable low-density sampling and instead focus on generating samples with high desired rewards. As shown in Table 1, SDG consistently outperforms its variant without the Stein correction, highlighting the importance of correcting posterior samples—approximated via Tweedie's formula—before applying them to the diffusion guidance tasks. In label guidance, SDG generates sharper images that improve the classification accuracy metric, though this comes at the cost of a higher FID due to the dataset's low resolution. Furthermore, SDG surpasses the relevant baselines, including DPS (Chung et al., 2023) and LGD (Song et al., 2023), which rely on approximated posterior samples exclusively.

Table 2: Mean and standard deviation of novel hit ratio (%) over three runs; baseline results from Lee et al. (2023). $\alpha(t)$ and $\epsilon(t)$ control low-density levels and particle update rates, respectively.

Method	NOVEL HIT RATIO (%) ↑		
	fa7	5ht1b	jak2
HierVAE (Jin et al., 2020)	0.007 (±0.013)	0.507 (± 0.278)	0.227 (±0.127)
MORLD (Jeon & Kim, 2020)	0.007 (±0.013)	0.880 (± 0.735)	0.227 (±0.118)
FREED (Yang et al., 2021)	1.107 (±0.209)	10.187 (± 3.306)	4.520 (±0.673)
GDSS (Jo et al., 2022)	0.368 (±0.103)	4.667 (± 0.306)	1.167 (±0.281)
MOOD (Lee et al., 2023)	0.733 (±0.141)	18.673 (±0.423)	9.200 (±0.524)
SDG w/o Stein correction	0.299 (±0.094)	0.033 (±0.027)	0.000 (±0.000)
SDG ($\alpha(t) > 0, \epsilon(t) > 0$)	1.156 (±0.087)	22.690 (±0.341)	9.167 (±0.262)
SDG ($\alpha(t) = 0, \epsilon(t) > 0$)	0.915 (±0.031)	21.278 (±0.332)	8.312 (±0.541)
SDG ($\alpha(t) > 0, \epsilon(t) = 0$)	0.956 (±0.247)	21.722 (±0.275)	8.722 (±0.218)

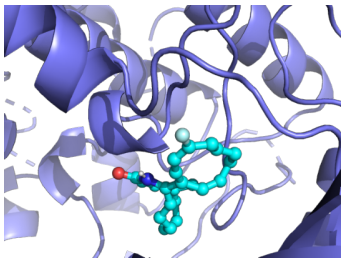

Figure 3: Example docking pose of a sampled ligand bound to the jak2 protein receptor.

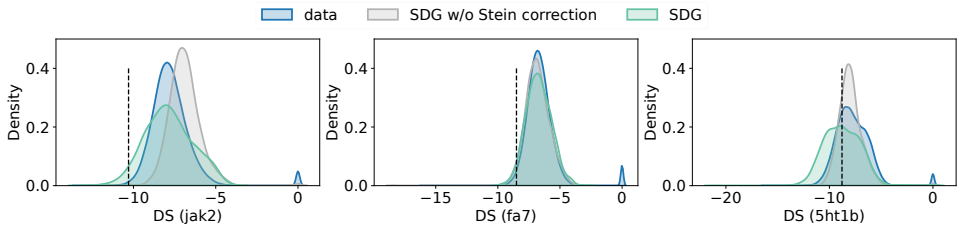

Figure 4: Distribution of docking scores (lower is better) for generated molecules of SDG with and without Stein correction. The black dashed lines indicate thresholds for identifying hit compounds.

4.2 LOW-DENSITY REWARD-BASED DIFFUSION GUIDANCE ON MOLECULES

To evaluate the hypothesis that sampling molecules from low-density regions can improve drug discovery, we apply SDG to the ligand–protein binding problem adapted from Lee et al. (2023). In this setting, generated molecules must satisfy four hit conditions: (1) Tanimoto similarity (**SIM**) with the closest training sample in ZINC250k (Irwin et al., 2012) is below 0.4 to ensure novelty (**Nov.**); (2) synthetic accessibility score (**SA**) is below 5, indicating ease of molecular synthesis; (3) drug-likeness score (**QED**) exceeds 0.5; and (4) docking score (**DS**) is lower than the median DS of known actives. We evaluate this task on three protein targets: **fa7** (Coagulation Factor VII), **5ht1b** (5-hydroxytryptamine receptor 1B), and **jak2** (Tyrosine-protein kinase JAK2). Detailed task descriptions, evaluation metrics, and baselines are provided in Appendix C.

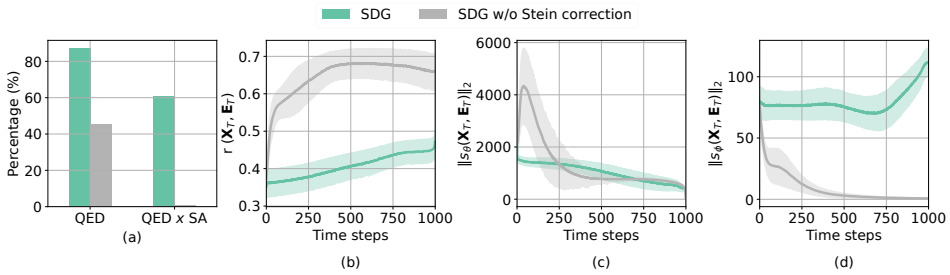

Figure 5: Temporal sampling dynamics of SDG for the jak2 protein. (a) Percentage of molecules meeting QED and SA hit criteria; (b) Rewards of posterior samples $r(\mathbf{X}_T, \mathbf{E}_T)$; (c) Frobenius norm of node posterior scores $s_\theta(\mathbf{X}_T, \mathbf{E}_T)$; (d) Frobenius norm of edge posterior scores $s_\phi(\mathbf{X}_T, \mathbf{E}_T)$.

4.2.1 Sampling novel hit molecules

As shown in Table 2, SDG without the Stein correction performs poorly, indicating that Tweedie's formula alone provides a biased and unreliable approximation of diffusion posteriors. Samples from these approximate posteriors fail to offer meaningful guidance, particularly in low-density regions where score-based models are least accurate due to limited training support. In contrast, SDG substantially enhances guidance by leveraging Stein-corrected samples, leading to improvements of several orders of magnitude. This not only validates the theoretical motivation for the Stein correction but also demonstrates consistent empirical gains. Moreover, SDG further boosts the performance of the pretrained GDSS model and its classifier-guided variant, MOOD, on two protein targets. Compared to non-diffusion methods, SDG generates more hit compounds across the target proteins. Figure 4 compares the distributions of docking scores. Without the Stein correction, SDG fails to align with the data distributions, resulting in poor guidance and fewer promising candidates. In contrast, SDG effectively regularizes sampling in low-density regions, shifting the docking score distributions toward the desired range and enabling the generation of more potential hit compounds.

Table 2 also shows that optimizing solely for rewards ($\alpha(t) = 0$) leads SDG to perform suboptimally, generating fewer novel hit molecules compared to the full setting ($\alpha(t) > 0$, $\epsilon(t) > 0$). This further supports the clear benefit of targeting low-density regions in molecular generation. SDG incorporates a novel Stein correction, which is provably more robust than the Langevin variant ($\epsilon(t) = 0$) of Corollary 3.5. These performance gains stem from the non-trivial interaction forces between particles, previously leveraged to increase sample diversity. Unlike Corso et al. (2024), however, SDG does not suffer from diversity issues, as most generated molecules remain unique (Figure 6). We visualize the docking pose of a sampled ligand on the jak2 target protein in Figure 3.

4.2.2 Reward overestimation and off-manifold sampling

In many applications, true (genuine) rewards are computed by non-differentiable oracle functions, which cannot be directly used in training-free diffusion guidance methods. Reward models and classifiers are trained to learn these genuine rewards and produce approximate (nominal) rewards, serving as differentiable proxies for diffusion models. However, due to the finite number of training samples, reward models tend to provide reliable signals only within the training data support. This limitation becomes more severe in low-density sampling problems. Uehara et al. (2024) first formulated this issue and proposed an entropy-regularized control approach, which is equivalent to solving a costly stochastic optimal control problem. In our case study, the absence of regularization leads the reward models to produce highly unreliable estimates of the true rewards during the sampling process (Figure 5.b), termed as *reward overestimation*. Moreover, this leads to the generation of

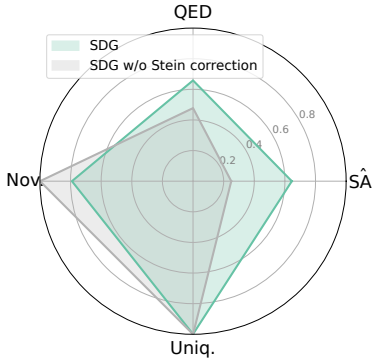

Figure 6: Multiple sampling objectives on jak2; SÂ denotes normalized synthetic accessibility (SA) scores.

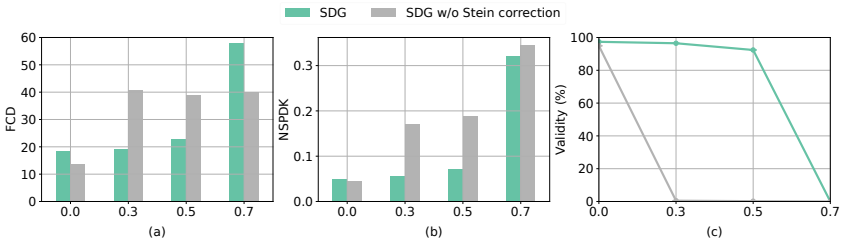

Figure 7: Ablation results under different low-density levels ($\alpha_{\max}$) for jak2. Chemical distance FCD (a) and structural distance NSPDK (b) compared to the test set; (c) Validity of generated molecules.

novel molecular structures that lack drug-likeness and are difficult to synthesize (Figures 5.a and 6). Additionally, the model scores vanish and are misdirected away from generative manifolds (Figure 5.c and 5.d). Further details on the sampling score dynamics are provided in Appendix C.3.3. In contrast, the Stein correction enables SDG to act as an efficient regularization mechanism that (i) improves genuine reward estimation accuracy (Figure 5.b), (ii) promotes the generation of more realistic and synthesizable molecules (Figure 5.a and Figure 6), and (iii) preserves sampling within the generative manifold (Figures 5.c and 5.d), particularly under challenging low-density conditions.

4.2.3 Ablation studies

Performance under extreme low-density sampling settings. We assess the robustness of SDG under varying low-density sampling conditions by measuring the chemical (FCD) and structural (NSPDK) distances between generated and test molecules. As shown in Figures 7(a,b), removing the Stein correction causes SDG to deviate substantially from the data distributions, even at modest low-density levels ($\alpha_{\max} = 0.3$). This degradation is also reflected at the molecular level, where most generated molecules exhibit invalid valency (Figure 7.c). In contrast, with the Stein correction, SDG guides samples within generative manifolds while marginally increasing FCD and NSPDK, thereby enabling sampling from lower-density regions. Moreover, the posterior correction produces significantly more valid structures, even under extreme conditions ($\alpha_{\max} = 0.5$), which underscores the effectiveness of Stein-based regularization for robust low-density molecular sampling scenarios.

Table 3: Ablation results on the number of particles.

Receptor	Novel Hit Ratio (%) ↑		
	512	1024	3000
fa7	1.156 (±0.087)	1.044 (±0.063)	1.067 (±0.109)
5ht1b	22.690 (±0.341)	21.922 (± 0.461)	22.389 (±0.490)
jak2	9.078 (±0.278)	8.822 (±0.532)	9.167 (±0.262)

Effect of particle size on Stein correction. SDG relies on Stein variational inference, whose effectiveness depends on the number of particles. While theory guarantees better approximation with more particles (Liu & Wang, 2016), our empirical results reveal a saturation point beyond which additional particles can yield inconsistent performance gains. Table 3 presents an ablation study across varying particle sizes, showing that larger particle sizes do not necessarily lead to better results. These behaviors likely arise from the inherent instability of kernel-based updates in high-dimensional spaces under the standard SVGD framework (Liu & Wang, 2016), consistent with prior analyses in Zhang et al. (2020), and highlight the need for more robust particle update schemes.

5 Conclusion

In this work, we propose Stein Diffusion Guidance (SDG), a low-cost alternative to SOC-based methods for enhancing diffusion guidance in a training-free manner. By analyzing the existing biases of Tweedie-based approximate posteriors through the lens of SOC theory, we introduce a plug-and-play Stein correction that effectively mitigates these biases. Experiments on low-density molecular sampling and general image guidance tasks provide strong empirical support for our theoretical claims. While effective, SDG still inherits the limitations of the standard SVGD formulation. Future work could explore more stable SVGD variants to further improve SDG for low-density diffusion guidance in high-dimensional settings.

REFERENCES

Amr Alhossary, Stephanus Daniel Handoko, Yuguang Mu, and Chee-Keong Kwoh. Fast, accurate, and reliable molecular docking with quickvina 2. *Bioinformatics*, 31(13):2214–2216, 2015.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OnD9zGAGT0k.

Gabriele Corso, Yilun Xu, Valentin De Bortoli, Regina Barzilay, and Tommi S. Jaakkola. Particle guidance: non-i.i.d. diverse sampling with diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=KqbCvIFBY7.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

Anh-Dung Dinh, Daochang Liu, and Chang Xu. Representative guidance: Diffusion model sampling with coherence. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=gWgaypDBs8.

Carles Domingo-Enrich, Jiequn Han, Brandon Amos, Joan Bruna, and Ricky T. Q. Chen. Stochastic optimal control matching. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=wfU2CdgmWt.

Carles Domingo-Enrich, Michal Drozdzal, Brian Karrer, and Ricky T. Q. Chen. Adjoint matching: Fine-tuning flow and diffusion generative models with memoryless stochastic optimal control. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=xQBRrtQM8u.

Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

Yingqing Guo, Hui Yuan, Yukang Yang, Minshuo Chen, and Mengdi Wang. Gradient guidance for diffusion models: An optimization perspective. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=X1QeUYBXke.

Carsten Hartmann and Christof Schütte. Efficient rare event simulation by optimal nonequilibrium forcing. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(11):P11004, 2012.

Elad Hazan and Sham Kakade. Revisiting the polyak step size. *arXiv preprint arXiv:1905.00313*, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

John J Irwin, Teague Sterling, Michael M Mysinger, Erin S Bolstad, and Ryan G Coleman. Zinc: a free tool to discover chemistry for biology. *Journal of chemical information and modeling*, 52 (7):1757–1768, 2012.

Yazid Janati, Badr Moufad, Mehdi Abou El Qassime, Alain Durmus, Eric Moulines, and Jimmy Olsson. A mixture-based framework for guiding diffusion models. *arXiv preprint arXiv:2502.03332*, 2025.

Woosung Jeon and Dongsup Kim. Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors. *Scientific reports*, 10(1):22104, 2020.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *International conference on machine learning*, pp. 4839–4848. PMLR, 2020.

Jaehyeong Jo, Seul Lee, and Sung Ju Hwang. Score-based generative modeling of graphs via the system of stochastic differential equations. In *International conference on machine learning*, pp. 10362–10383. PMLR, 2022.

Hilbert J Kappen, Vicenç Gómez, and Manfred Opper. Optimal control as a graphical model inference problem. *Machine learning*, 87:159–182, 2012.

Seul Lee, Jaehyeong Jo, and Sung Ju Hwang. Exploring chemical space with score-based out-of-distribution generation. In *International Conference on Machine Learning*, pp. 18872–18892. PMLR, 2023.

Qiang Liu. Stein variational gradient descent as gradient flow. *Advances in neural information processing systems*, 30, 2017.

Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems*, 29, 2016.

Qiang Liu, Jason Lee, and Michael Jordan. A kernelized stein discrepancy for goodness-of-fit tests. In *International conference on machine learning*, pp. 276–284. PMLR, 2016.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Jiajun Ma, Tianyang Hu, Wenjia Wang, and Jiacheng Sun. Elucidating the design space of classifier-guided diffusion generation. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=9DXXMXnIGm.

Badr MOUFAD, Yazid Janati, Lisa Bedin, Alain Oliviero Durmus, randal douc, Eric Moulines, and Jimmy Olsson. Variational diffusion posterior sampling with midpoint guidance. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=6EUtjXAvmj.

Nikolas Nüsken and Lorenz Richter. Solving high-dimensional hamilton–jacobi–bellman pdes using neural networks: perspectives from the theory of controlled diffusions and measures on path space. *Partial differential equations and applications*, 2(4):48, 2021.

Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

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

Warren B Powell. From reinforcement learning to optimal control: A unified framework for sequential decisions. In *Handbook of Reinforcement Learning and Control*, pp. 29–74. Springer, 2021.

Herbert E Robbins. An empirical bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and basic theory*, pp. 388–394. Springer, 1992.

Litu Rout, Yujia Chen, Nataniel Ruiz, Abhishek Kumar, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. RB-modulation: Training-free stylization using reference-based modulation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=bnINPG5A32.

Vikash Sehwag, Caner Hazirbas, Albert Gordo, Firat Ozgenel, and Cristian Canton. Generating high fidelity data from low-density regions using diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11492–11501, 2022.

Yifei Shen, Xinyang Jiang, Yifan Yang, Yezhen Wang, Dongqi Han, and Dongsheng Li. Understanding and improving training-free loss-based diffusion guidance. *Advances in Neural Information Processing Systems*, 37:108974–109002, 2024.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *International Conference on Machine Learning*, pp. 32483–32498. PMLR, 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum likelihood training of score-based diffusion models. *Advances in neural information processing systems*, 34:1415–1428, 2021.

Belinda Tzen and Maxim Raginsky. Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit. *arXiv preprint arXiv:1905.09883*, 2019.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*, 2024.

Chenyu Wang, Masatoshi Uehara, Yichun He, Amy Wang, Avantika Lal, Tommi Jaakkola, Sergey Levine, Aviv Regev, Hanchen, and Tommaso Biancalani. Fine-tuning discrete diffusion models via reward optimization with applications to DNA and protein design. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=G328D1xt4W.

Soojung Yang, Doyeong Hwang, Seul Lee, Seongok Ryu, and Sung Ju Hwang. Hit and lead discovery with explorative rl and fragment-based molecule generation. *Advances in Neural Information Processing Systems*, 34:7924–7936, 2021.

Haotian Ye, Haowei Lin, Jiaqi Han, Minkai Xu, Sheng Liu, Yitao Liang, Jianzhu Ma, James Zou, and Stefano Ermon. TFG: Unified training-free guidance for diffusion models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=N8YbGX98vc.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 23174–23184, 2023.

Jianyi Zhang, Ruiyi Zhang, Lawrence Carin, and Changyou Chen. Stochastic particle-optimization sampling and the non-asymptotic convergence theory. In *International Conference on Artificial Intelligence and Statistics*, pp. 1877–1887. PMLR, 2020.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 586–595, 2018.

Algorithm 1 Stein Diffusion Guidance Algorithm

Input: Score model $s_\theta(\mathbf{x}_t)$, number of particles N, off-the-shelf classifier $r(\mathbf{x}_T)$, correction step size $\epsilon(t)$, low-density schedule $\alpha(t)$, guidance strength schedule $\beta(t)$, total steps T.
Output: Endpoint samples.
$\mathcal{D}_0 \leftarrow \left\{\mathbf{x}_0^i | \mathbf{x}_0^i \sim p_0, 1 \leqslant i \leqslant N\right\}$ {initial particles}
for t in $[0, T)$:
/ Back-and-forth Stein correction */*
  $\mathcal{D}_T \leftarrow \left\{\mathbf{x}_T^i | \mathbf{x}_T^i = \frac{\mathbf{x}_t^i + \gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t^i)}{\eta(t)}, \forall \mathbf{x}_t^i \in \mathcal{D}_t\right\}$ {Backward mapping}
  $\mathbf{x}_T^i = \mathbf{x}_T^i + \epsilon(t)\frac{1}{N}\sum_{\mathbf{x}_t^j, \mathbf{x}_T^j \in \mathcal{D}_t \bigcup \mathcal{D}_T}\left((s_\theta(\mathbf{x}_T^j) - \eta(t)s_\theta(\mathbf{x}_t^j))k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j} k(\mathbf{x}_T^i, \mathbf{x}_T^j)\right)$
  $\mathcal{D}_t \leftarrow \left\{\mathbf{x}_t^i | \mathbf{x}_t^i \sim \mathcal{N}(\eta(t)\mathbf{x}_T^i, \gamma^2(t)I), \forall \mathbf{x}_T^i \in \mathcal{D}_T\right\}$ {Forward mapping}
/ Low-density reward-based diffusion guidance */*
  $\mathbf{x}_T^i = \frac{\mathbf{x}_t^i + \gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t^i)}{\eta(t)}$
  $\mathbf{x}_{t+1}^i = \mathbf{x}_t^i - \mathbf{b}(\mathbf{x}_t^i, t) + \sigma(t)^2\left((1-\alpha(t))s_\theta(\mathbf{x}_t^i) + \beta(t)\nabla_{\mathbf{x}_t^i} r(\mathbf{x}_T^i)\right) + \sigma(t)\mathbf{z}, \;\; \mathbf{z} \sim \mathcal{N}(0, I)$
return: $\left\{\mathbf{x}_T^i, \; 1 \leqslant i \leqslant N\right\}$

A Broader impact statements

Our work presents a novel, generic approach to training-free diffusion guidance by connecting with Stein variational inference and stochastic optimal control theory. The target application is for low-density molecular sampling, with the potential to advance drug discovery by finding effective drug candidates for cancer treatments. However, in the wrong hands, it could be misused to illicitly design harmful or addictive substances. Ensuring responsible usage is, therefore, critical to maximizing its positive societal impact.

B Theoretical proofs

Proposition 3.1 *Consider the SOC problem in Equation 2 with the novel functional cost $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ defined in Equation 3. By Lemma 2.2, the marginal density $p_t^{\mathbf{u}}(\mathbf{x}_t)$, the value function $V(\mathbf{x}, t)$, and the optimal control $\mathbf{u}^*(\mathbf{x}, t)$ of the controlled-reverse SDE under $\mathbb{P}^{\mathbf{u}}$ are given as*

$$p_t^{\mathbf{u}}(\mathbf{x}_t) = p_t^{1-\alpha(t)}(\mathbf{x}_t)\exp(\beta(t)r(\mathbf{x}_T))$$
$$\mathbf{u}^*(\mathbf{x}, t) = \sigma(t)\nabla_{\mathbf{x}}\log\mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})}|\mathbf{x}_t = \mathbf{x}\right] \;\text{ and }\; V(\mathbf{x}, t) = -\log\mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})}|\mathbf{x}_t = \mathbf{x}\right]$$

Proof. For low-density sampling, we introduce a state cost that penalizes the density of the current state, expressed as $f(\mathbf{x}_s, s) = \alpha(s)\log p_s(\mathbf{x}_s)\,\delta(s-t)$, where $\alpha(s)$ defines a low-density annealing schedule. To maximize the reward of generated samples, we define a terminal cost as $g(\mathbf{x}_T) = -\beta(t)r(\mathbf{x}_T)$, where $r(\cdot)$ is the reward function and $\beta(t)$ defines a guidance-strength schedule. These two terms form a novel functional cost function $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ (Equation 3). By Lemma 2.2, we obtain the value function for this cost function.

$$\begin{aligned} V(\mathbf{x}, t) &= -\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\int_t^T \alpha(s)\log p_s(\mathbf{x}_s)\,\delta(s-t)ds + \beta(t)r(\mathbf{x}_T)\right)|\mathbf{x}_t = \mathbf{x}\right] \\ &= -\log\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\alpha(t)\log p_t(\mathbf{x}) + \beta(t)r(\mathbf{x}_T)\right)|\mathbf{x}_t = \mathbf{x}\right] \\ &= -\log\mathbb{E}_{\mathbb{P}}\left[p_t^{-\alpha(t)}(\mathbf{x})\exp\left(\beta(t)r(\mathbf{x}_T)\right)|\mathbf{x}_t = \mathbf{x}\right] \end{aligned} \tag{7}$$

And, the optimal control has the form.

$$\mathbf{u}^*(\mathbf{x},t) = \sigma(t)\nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[p_t^{-\alpha(t)}(\mathbf{x}) \exp\left(\beta(t) r(\mathbf{x}_T)\right) | \mathbf{x}_t = \mathbf{x}\right] \tag{8}$$

Substituting the optimal control into the stochastic optimal control problem (Equation 2) yields the controlled SDE under $\mathbb{P}^{\mathbf{u}}$.

$$\begin{aligned}
d\mathbf{x}_t^{\mathbf{u}} &= \left(-\mathbf{b}(\mathbf{x}_t^{\mathbf{u}},t) + \sigma(t)^2 \nabla_{\mathbf{x}_t^{\mathbf{u}}} \log p_t(\mathbf{x}_t^{\mathbf{u}}) + \sigma(t)\mathbf{u}(\mathbf{x}_t^{\mathbf{u}},t)\right) dt + \sigma(t) d\mathbf{w} \\
&= \left(-\mathbf{b}(\mathbf{x}_t^{\mathbf{u}},t) + \sigma(t)^2 \nabla_{\mathbf{x}_t^{\mathbf{u}}} \log p_t(\mathbf{x}_t^{\mathbf{u}}) + \sigma(t)^2 \nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[p_t^{-\alpha(t)}(\mathbf{x}) \exp\left(\beta(t) r(\mathbf{x}_T)\right) | \mathbf{x}_t^{\mathbf{u}} = \mathbf{x}\right]\right) dt + \sigma(t) d\mathbf{w} \\
&= \left(-\mathbf{b}(\mathbf{x}_t^{\mathbf{u}},t) + \sigma(t)^2 \nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[p_t^{1-\alpha(t)}(\mathbf{x}) \exp\left(\beta(t) r(\mathbf{x}_T)\right) | \mathbf{x}_t^{\mathbf{u}} = \mathbf{x}\right]\right) dt + \sigma(t) d\mathbf{w} \\
&= \left(-\mathbf{b}(\mathbf{x}_t^{\mathbf{u}},t) + \sigma(t)^2 \nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[p_t^{\mathbf{u}}(\mathbf{x}) | \mathbf{x}_t^{\mathbf{u}} = \mathbf{x}\right]\right) dt + \sigma(t) d\mathbf{w}
\end{aligned} \tag{9}$$

where $p_t^{\mathbf{u}}(\mathbf{x}_t)$ denotes the annealed, reward-guided marginal density under $\mathbb{P}^{\mathbf{u}}$:

$$p_t^{\mathbf{u}}(\mathbf{x}_t) = p_t^{1-\alpha(t)}(\mathbf{x}_t) \exp(\beta(t) r(\mathbf{x}_T)) \tag{10}$$

Substituting this expression back into the value function and optimal control, we conclude the proof.

$$\mathbf{u}^*(\mathbf{x},t) = \sigma(t)\nabla_{\mathbf{x}} \log \mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \quad \text{and} \quad V(\mathbf{x},t) = -\log \mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \tag{11}$$

Proposition 3.2 *Let $q \in \mathcal{Q}$ be any traceable family of proposal distributions. Then, the value function in Proposition 3.1 admits the following upper bound:*

$$\begin{aligned}
V(\mathbf{x},t) &\le \bar{V}(\mathbf{x},t,q) \\
&= -\mathbb{E}_{\mathbf{x}_T \sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[\log\left(\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})}\right) | \mathbf{x}_t = \mathbf{x}\right] + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t) \| p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t = \mathbf{x}} \\
&= \alpha(t) \log p_t(\mathbf{x}) - \beta(t) \mathbb{E}_{\mathbf{x}_T \sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[r(\mathbf{x}_T) | \mathbf{x}_t = \mathbf{x}\right] \\
&\qquad + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t) \| p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t = \mathbf{x}}
\end{aligned}$$

Proof. We begin by rewriting the value function in terms of the diffusion posterior $p(\mathbf{x}_T \mid \mathbf{x}_t)$, which denotes the terminal distribution evolved from $\mathbf{x}_t$ under the uncontrolled process $\mathbb{P}$.

$$\begin{aligned}
V(\mathbf{x},t) &= -\log \mathbb{E}_{\mathbb{P}}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \\
&= -\log \mathbb{E}_{p(\mathbf{x}_T|\mathbf{x})}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \\
&= -\log \int \frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} p(\mathbf{x}_T|\mathbf{x}) d\mathbf{x}_T, \quad \textit{note: } p_t^{\mathbf{u}}(\mathbf{x}) \textit{ depends implicitly on } \mathbf{x}_T \\
&= -\log \int \frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} \frac{p(\mathbf{x}_T|\mathbf{x})}{q(\mathbf{x}_T|\mathbf{x})} q(\mathbf{x}_T|\mathbf{x}) d\mathbf{x}_T, \quad \textit{where } q(\mathbf{x}_T|\mathbf{x}) \textit{ is a traceable simpler distribution.} \\
&= -\log \mathbb{E}_{q(\mathbf{x}_T|\mathbf{x})}\left[\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})} \frac{p(\mathbf{x}_T|\mathbf{x})}{q(\mathbf{x}_T|\mathbf{x})} | \mathbf{x}_t = \mathbf{x}\right] \\
&\le -\mathbb{E}_{\mathbf{x}_T \sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[\log\left(\frac{p_t^{\mathbf{u}}(\mathbf{x})}{p_t(\mathbf{x})}\right) | \mathbf{x}_t = \mathbf{x}\right] + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t) \| p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t = \mathbf{x}}, \textit{Jensen's inequality} \\
&= \alpha(t) \log p_t(\mathbf{x}) - \beta(t) \mathbb{E}_{\mathbf{x}_T \sim q_{T|t}(\mathbf{x}_T|\mathbf{x})}\left[r(\mathbf{x}_T) | \mathbf{x}_t = \mathbf{x}\right] + D_{KL}\left(q(\mathbf{x}_T|\mathbf{x}_t) \| p(\mathbf{x}_T|\mathbf{x}_t)\right)|_{\mathbf{x}_t = \mathbf{x}} \\
&= \bar{V}(\mathbf{x},t,q)
\end{aligned} \tag{12}$$

Lemma 3.3 *For $\mathbf{x}_T \sim p(\mathbf{x}_T|\mathbf{x}_t)$, the posterior score admits the following approximation in terms of the score model $\mathbf{s}_{\boldsymbol{\theta}}(\cdot)$:*

$$\nabla_{\mathbf{x}_T} \log p(\mathbf{x}_T|\mathbf{x}_t) \approx \mathbf{s}_\theta(\mathbf{x}_T) - \eta(t)\mathbf{s}_\theta(\mathbf{x}_t)$$

Proof. We commence by expressing the marginal data score as the expectation of the conditional data score over the posterior distribution.

$$\begin{aligned}
\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) &= \nabla_{\mathbf{x}_t} \log \int p(\mathbf{x}_t, \mathbf{x}_T) d\mathbf{x}_T \\
&= \nabla_{\mathbf{x}_t} \log \int p(\mathbf{x}_t \mid \mathbf{x}_T) p(\mathbf{x}_T) d\mathbf{x}_T \\
&= \frac{1}{p(\mathbf{x}_t)} \nabla_{\mathbf{x}_t} \int p(\mathbf{x}_t \mid \mathbf{x}_T) p(\mathbf{x}_T) d\mathbf{x}_T, \ \textit{Identity's rule} \\
&= \frac{1}{p(\mathbf{x}_t)} \int \nabla_{\mathbf{x}_t} p(\mathbf{x}_t \mid \mathbf{x}_T) p(\mathbf{x}_T) d\mathbf{x}_T \\
&= \frac{1}{p(\mathbf{x}_t)} \int p(\mathbf{x}_T, \mathbf{x}_t) \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_T) d\mathbf{x}_T \\
&= \frac{1}{p(\mathbf{x}_t)} \int p(\mathbf{x}_t \mid \mathbf{x}_T) \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_T) p(\mathbf{x}_T) d\mathbf{x}_T, \ \textit{Identity's rule} \\
&= \int p(\mathbf{x}_T \mid \mathbf{x}_t) \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t \mid \mathbf{x}_T) d\mathbf{x}_T \\
&= E_{p(\mathbf{x}_T|\mathbf{x}_t)}[\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_T)] \qquad (13)
\end{aligned}$$

We leverage the identity, $\nabla_x \log f(x) = \frac{1}{f(x)} \nabla_x f(x)$, in two critical steps. Applying a one-sample Monte Carlo estimation of the expectation, we obtain the following approximation.

$$s_\theta(\mathbf{x}_t) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_T), \qquad \mathbf{x}_T \sim p(\mathbf{x}_T|\mathbf{x}_t) \qquad (14)$$

Assuming the noise kernel with the form $p_{t|T}(\mathbf{x}_t|\mathbf{x}_T) = \mathcal{N}(\eta(t)\mathbf{x}_T, \gamma^2(t)I)$, the model score can be expressed as:

$$s_\theta(\mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_T) = -\frac{\mathbf{x}_t - \eta(t)\mathbf{x}_T}{\gamma^2(t)} = -\frac{\mathbf{z}}{\gamma(t)}, \qquad \mathbf{z} \sim \mathcal{N}(0, I) \qquad (15)$$

The last equality results from $\mathbf{x}_t = \eta(t)\mathbf{x}_T + \gamma(t)\mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, I)$. Moreover, the posterior score can be decomposed as:

$$\begin{aligned}
\nabla_{\mathbf{x}_T} \log p(\mathbf{x}_T|\mathbf{x}_t) &= \nabla_{\mathbf{x}_T} \log \frac{p(\mathbf{x}_t|\mathbf{x}_T)p(\mathbf{x}_T)}{p(\mathbf{x}_t)} \\
&= \nabla_{\mathbf{x}_T} \log p(\mathbf{x}_t|\mathbf{x}_T) + \nabla_{\mathbf{x}_T} \log p(\mathbf{x}_T) \\
&= \eta(t) \frac{\mathbf{x}_t - \eta(t)\mathbf{x}_T}{\gamma^2(t)} + s_\theta(\mathbf{x}_T) \\
&= \eta(t) \frac{\mathbf{z}}{\gamma(t)} + s_\theta(\mathbf{x}_T) \\
&\approx s_\theta(\mathbf{x}_T) - \eta(t) s_\theta(\mathbf{x}_t) \qquad (16)
\end{aligned}$$

We apply the model score to the last approximation, and $\mathbf{x}_T \sim p(\mathbf{x}_T|\mathbf{x}_t)$, we conclude the proof.

Proposition 3.4 *Consider the low-density reward-based cost functional $\widetilde{J}(\mathbf{u}, \mathbf{x}, t)$ and its upper bound value function $\bar{V}(\mathbf{x}, t, q)$. Let $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ denote the proposal posterior initialized via Tweedie's formula, and let $q^{\epsilon}_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ denote the updated posterior obtained after applying the back-and-forth Stein correction with step size $\epsilon(t)$. Then, the optimal control $\bar{\mathbf{u}}^*(\mathbf{x}, t)$ for the surrogate value function $\bar{V}(\mathbf{x}, t, q)$ decomposed as*

$$\frac{\bar{\mathbf{u}}^*(\mathbf{x}_t^i, t)}{\sigma(t)} = \underbrace{-\alpha(t)\mathbf{s}_\theta\left(\mathbf{x}_t^i\right) + \beta(t)\nabla_{\mathbf{x}_t^i}\mathbb{E}_{\mathbf{x}_T^i \sim q^{\epsilon}_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[r(\mathbf{x}_T^i)\right]}_{\textbf{Low-density reward-based guidance on } \mathcal{M}_t}$$
$$\oplus \quad \underbrace{\mathbb{E}_{\mathbf{x}_T^j \sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[\left(\mathbf{s}_\theta(\mathbf{x}_T^j) - \eta(t)\mathbf{s}_\theta(\mathbf{x}_t^j)\right)k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j}k(\mathbf{x}_T^i, \mathbf{x}_T^j)\right]}_{\textbf{Stein diffusion posterior correction on } \mathcal{M}_T}$$

Proof. By Lemma 2.2, we obtain the optimal control of the surrogate value function.

$$\begin{aligned}
&\frac{\bar{\mathbf{u}}^*(\mathbf{x}_t^i, t)}{\sigma(t)} = -\nabla_{\mathbf{x}_t^i}\bar{V}(\mathbf{x}_t^i, t, q) \\
&= \underbrace{-\nabla_{\mathbf{x}_t^i} D_{KL}\left(q(\mathbf{x}_T^i|\mathbf{x}_t^i)\|p(\mathbf{x}_T^i|\mathbf{x}_t^i)\right)}_{\textbf{I}} + \underbrace{-\alpha(t)\mathbf{s}_\theta\left(\mathbf{x}_t^i\right) + \beta(t)\nabla_{\mathbf{x}_t^i}\mathbb{E}_{\mathbf{x}_T^i \sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[r(\mathbf{x}_T^i)\right]}_{\textbf{II}} \quad (17)
\end{aligned}$$

The first control component (I) guides posterior samples in the direction that minimizes the KL divergence between the approximate and true posteriors. The second control component (II) enables low-density sampling and reward/classifier-based diffusion guidance. Most existing training-free diffusion guidance methods utilize solely the second control component while ignoring the first one, which thus does not guarantee sampling from the true posterior. In this work, we propose Stein Diffusion Guidance (SDG), which incorporates the guidance control from both components. Let's consider a proposal posterior $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$, whose mean value is estimated via Tweedie's formula $(\mathbf{x}_t + \gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t, t))/\eta(t)$. Below, we present the analytical form of each control component.

The KL divergence control (I): Since the true posterior $p(\mathbf{x}_T^i|\mathbf{x}_t^i)$ does not admit a closed-form expression, we can not compute analytically the KL divergence and its gradient. To address this, we leverage the Stein variational inference, a nonparametric approach that identifies the steepest gradient direction to minimize the KL divergence. Assuming a batch of N particles at the t^{th} reverse diffusion timestep $\mathcal{D}_t \leftarrow \{\mathbf{x}_t^i\}_{i=0}^N$, we apply Lemma 2.1 to obtain the KSD's direction minimizing $D_{KL}\left(q(\mathbf{x}_T^i|\mathbf{x}_t^i)\|p(\mathbf{x}_T^i|\mathbf{x}_t^i)\right)$:

$$\phi^*(\mathbf{x}_t^i) = \mathbb{E}_{\mathbf{x}_T^j \sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}[\nabla_{\mathbf{x}_t^j}\log p(\mathbf{x}_T^j|\mathbf{x}_t^j)k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_t^j}k(\mathbf{x}_T^i, \mathbf{x}_T^j)]$$

Computing this optimal direction requires numerous Jacobian-vector product evaluations, e.g, $\nabla_{\mathbf{x}_T^j}\log p(\mathbf{x}_T^j|\mathbf{x}_t^j)\frac{\partial \mathbf{x}_T^j}{\partial \mathbf{x}_t^j}$ and $\nabla_{\mathbf{x}_T^j}k(\mathbf{x}_T^i, \mathbf{x}_T^j)\frac{\partial \mathbf{x}_T^j}{\partial \mathbf{x}_t^j}$, which are computationally expensive as the number of particles N increases. To alleviate this computational burden, we propose *a back-and-forth Stein correction*: (**i**) Apply Tweedie's formula to map backward the particles $\mathcal{D}_t \leftarrow \{\mathbf{x}_t^i\}_{i=0}^N$ on $\mathcal{M}_t$ to $\mathcal{D}_T \leftarrow \left\{\mathbf{x}_T^i|\mathbf{x}_T^i = \frac{\mathbf{x}_t^i + \gamma^2(t)\mathbf{s}_\theta(\mathbf{x}_t^i)}{\eta(t)}, \forall \mathbf{x}_t^i \in \mathcal{D}_t\right\}$ on $\mathcal{M}_T$, which represents the initial proposal posterior $q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$; (**ii**) Apply the Stein correction on the particles of $\mathcal{D}_T$, which results in the corrected posterior $q^{\epsilon}_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$; (**iii**) Apply the perturbation kernel $p_{t|T}(\mathbf{x}_t|\mathbf{x}_T)$ to map forward the Stein-corrected particles $\mathcal{D}_T$ on $\mathcal{M}_T$ to $\mathcal{D}_t \leftarrow \left\{\mathbf{x}_i^t|\mathbf{x}_i^t \sim \mathcal{N}(\eta(t)\mathbf{x}_T^i, \gamma^2(t)I), \forall \mathbf{x}_T^i \in \mathcal{D}_T\right\}$ on $\mathcal{M}_t$. In the second step, the Stein correction applies a particle-based transform on each particle of $\mathcal{D}_T$, which follows the KSD's direction, given as:

$$\phi^*(\mathbf{x}_T^i) = \mathbb{E}_{\mathbf{x}_T^j \sim q_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}[\nabla_{\mathbf{x}_T^j}\log p(\mathbf{x}_T^j|\mathbf{x}_t^j)k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j}k(\mathbf{x}_T^i, \mathbf{x}_T^j)] \quad (18)$$

We observe that the Stein correction on $\mathcal{M}_T$ does not involve any evaluations of Jacobian-vector products, which achieves more memory efficiency during inference. By replacing the posterior score from Lemma 3.3, we obtain the KSD's direction with a closed form.

$$\phi^*(\mathbf{x}_T^i) = \mathbb{E}_{\mathbf{x}_T^j \sim q_{T|t}(\mathbf{x}_T|\mathcal{D}_t)}[(-\eta(t)\mathbf{s}_\theta(\mathbf{x}_t^j) + \mathbf{s}_\theta(\mathbf{x}_T^j))k(\mathbf{x}_T^i, \mathbf{x}_T^j) + \nabla_{\mathbf{x}_T^j} k(\mathbf{x}_T^i, \mathbf{x}_T^j)] \tag{19}$$

The particle update with a stepsize $\epsilon(t)$ can be taken as:

$$\mathbf{x}_T^i = \mathbf{x}_T^i + \epsilon(t) * \phi^*(\mathbf{x}_T^i) \tag{20}$$

The low-density reward guidance control (II): We use the Stein-corrected posterior $q^{\epsilon}_{T|t}(\mathbf{x}_T|\mathbf{x}_t)$ to guide samples toward low-density regions with high rewards or desired properties. The optimal control for this task can be written in an analytical form.

$$-\alpha(t)\mathbf{s}_\theta\left(\mathbf{x}_t^i\right) + \beta(t)\nabla_{\mathbf{x}_t^i}\mathbb{E}_{\mathbf{x}_T^i \sim q^{\epsilon}_{T|t}(\mathbf{x}_T|\mathbf{x}_t)}\left[r(\mathbf{x}_T^i)\right] \tag{21}$$

Where $\mathbf{x}_t^i$ is the noised sample on $\mathcal{M}_t$, resulting from applying the forward kernel on its Stein-corrected version $\mathbf{x}_T^i$ on $\mathcal{M}_T$.

By concatenating the two control components, we conclude the proof.

Corollary 3.5 *Let the correction stepsize be set to zero, $\epsilon(t) = 0$ for all t, then the back-and-forth Stein correction reduces to the Langevin correction with stepsize $\gamma^2(t)$ and noise scaled by $\sqrt{2}$:*

$$\mathbf{x}_t \leftarrow \mathbf{x}_t + \gamma^2(t)s_\theta(\mathbf{x}_t) + \gamma(t)\mathbf{z}, \qquad \mathbf{z} \sim \mathcal{N}(0, I)$$

Proof. By setting $\epsilon(t) = 0$ for all t, from Algorithm 1, we have the back-and-forth Stein correction mechanism with a following analytical form:

$$\begin{aligned}\mathbf{x}_t &\leftarrow \eta(t)\frac{\mathbf{x}_t + \gamma^2(t)s_\theta(\mathbf{x}_t)}{\eta(t)} + \gamma(t)\mathbf{z} \\ &= \mathbf{x}_t + \gamma^2(t)s_\theta(\mathbf{x}_t) + \gamma(t)\mathbf{z}, \qquad \mathbf{z} \sim \mathcal{N}(0, I)\end{aligned} \tag{22}$$

This corresponds to the Langevin correction from Song et al. (2020b), with a step size of $\gamma^2(t)$ and the noise term scaled down by a factor of $\sqrt{2}$. As a result, the back-and-forth Stein correction generalizes the Langevin correction as a special case.

C Molecular sampling in low-density regions

C.1 Sampling molecular graph permutation-invariant distributions

Our target application of Stein Diffusion Guidance is to enable sampling molecular graphs with desired rewards in low-density regions in a training-free diffusion guidance manner. We adapt the score-based generative framework from Jo et al. (2022) for modeling molecular graph distributions. Given a graph representation $\mathcal{G} = (\mathbf{X}, \mathbf{E})$, with $\mathbf{X}$ and $\mathbf{E}$ denoting the node and edge feature matrices, respectively, the authors introduce a *system* of SDEs to capture molecular graph distributions, whose reverse/generative process can be derived as:

$$\begin{cases} \mathbf{X}_{t+1} = \mathbf{X}_t - \mathbf{b}_{\mathbf{X}}(\mathbf{X}_t, t) + \sigma_{\mathbf{X}}(t)^2 s_\theta(\mathbf{X}_t, \mathbf{E}_t) + \sigma_{\mathbf{X}}(t)\mathbf{Z}, & \mathbf{Z} \sim \mathcal{N}(0, I) \\ \mathbf{E}_{t+1} = \mathbf{E}_t - \mathbf{b}_{\mathbf{E}}(\mathbf{E}_t, t) + \sigma_{\mathbf{E}}(t)^2 s_\phi(\mathbf{X}_t, \mathbf{E}_t) + \sigma_{\mathbf{E}}(t)\mathbf{Z}, & \mathbf{Z} \sim \mathcal{N}(0, I) \end{cases} \tag{23}$$

The authors utilize two score networks to approximate conditional data scores: $s_\theta(\mathbf{X}_t, \mathbf{E}_t) \approx \nabla_{\mathbf{X}_t} \log p(\mathbf{X}_t, \mathbf{E}_t)$ and $s_\phi(\mathbf{X}_t, \mathbf{E}_t) \approx \nabla_{\mathbf{E}_t} \log p(\mathbf{X}_t, \mathbf{E}_t)$. In addition, $s_{\theta,\phi}(\mathbf{X}_t, \mathbf{E}_t)$ are permutation-equivariant models that respect the inherent symmetry of graph data. This system of coupled SDEs must be solved simultaneously in order to sample graph distributions. Building on this unconditional sampling foundation, we extend Stein Diffusion Guidance to sample molecular graphs with desired properties in low-density regions. From Algorithm 1, SDG first corrects posterior molecular graph samples on the manifold $\mathcal{M}_T$:

$$\begin{cases} \mathbf{X}_T^i = \mathbf{X}_T^i + \epsilon_{\mathbf{X}}(t) \frac{1}{N} \sum_{\mathbf{X}_t^j, \mathbf{X}_T^j \in \mathcal{D}_t \bigcup \mathcal{D}_T} \Big((s_\theta(\mathbf{X}_T^j, \mathbf{E}_T^j) - \eta_{\mathbf{X}}(t) s_\theta(\mathbf{X}_t^j, \mathbf{E}_t^j)) k(\mathbf{X}_T^i, \mathbf{X}_T^j) + \nabla_{\mathbf{X}_T^j} k(\mathbf{X}_T^i, \mathbf{X}_T^j) \Big) \\ \mathbf{E}_T^i = \mathbf{E}_T^i + \epsilon_{\mathbf{E}}(t) \frac{1}{N} \sum_{\mathbf{E}_t^j, \mathbf{E}_T^j \in \mathcal{D}_t \bigcup \mathcal{D}_T} \Big((s_\phi(\mathbf{X}_T^j, \mathbf{E}_T^j) - \eta_{\mathbf{E}}(t) s_\phi(\mathbf{X}_t^j, \mathbf{E}_t^j)) k(\mathbf{E}_T^i, \mathbf{E}_T^j) + \nabla_{\mathbf{E}_T^j} k(\mathbf{E}_T^i, \mathbf{E}_T^j) \Big) \end{cases} \tag{24}$$

And then utilizing the Stein-corrected posterior samples to perform low-density diffusion guidance with off-the-shelf molecular property predictors; we refer to property predictors as classifiers.

$$\begin{cases} \mathbf{X}_{t+1}^i = \mathbf{X}_t^i - \mathbf{b}_{\mathbf{X}}(\mathbf{X}_t^i, t) + \sigma_{\mathbf{X}}(t)^2 \Big((1 - \alpha_{\mathbf{X}}(t)) s_\theta(\mathbf{X}_t^i, \mathbf{E}_t^i) + \beta_{\mathbf{X}}(t) \nabla_{\mathbf{X}_t^i} r(\mathbf{X}_T^i, \mathbf{E}_T^i) \Big) + \sigma_{\mathbf{X}}(t) \mathbf{Z} \\ \mathbf{E}_{t+1}^i = \mathbf{E}_t^i - \mathbf{b}_{\mathbf{E}}(\mathbf{E}_t^i, t) + \sigma_{\mathbf{E}}(t)^2 \Big((1 - \alpha_{\mathbf{E}}(t)) s_\phi(\mathbf{X}_t^i, \mathbf{E}_t^i) + \beta_{\mathbf{E}}(t) \nabla_{\mathbf{E}_t^i} r(\mathbf{X}_T^i, \mathbf{E}_T^i) \Big) + \sigma_{\mathbf{E}}(t) \mathbf{Z} \end{cases} \tag{25}$$

We obtain initial posterior samples via Tweedie's formula, given as $\mathbf{X}_T^i = (\mathbf{X}_t^i + \gamma_{\mathbf{X}}(t) s_\theta(\mathbf{X}_t^i, \mathbf{E}_t^i))/\eta_{\mathbf{X}}(t)$ and $\mathbf{E}_T^i = (\mathbf{E}_t^i + \gamma_{\mathbf{E}}(t) s_\phi(\mathbf{X}_t^i, \mathbf{E}_t^i))/\eta_{\mathbf{E}}(t)$.

Based on the primary work (Jo et al., 2022), Lee et al. (2023) further propose a standard classifier-based diffusion guidance for conditional molecular sampling in out-of-distribution settings. In experiments, we use the same pretrained models and settings as Lee et al. (2023). Concretely, we model the node component using a Variance Preserving SDE (VPSDE) and the edge component using a Variance Exploding SDE (VESDE). In sampling, we adopt the predictor-corrector scheme, with the reverse SDE as the predictor and annealed Langevin dynamics as the corrector. As reported in Jo et al. (2022) (Table 12), the pretrained models tend to sample molecules with very low validity when using either the predictor or corrector framework alone.

C.2 Experimental setup

Datasets We benchmark SDG on molecular generation tasks aimed at discovering novel ligands with strong binding affinity to specific protein targets. Following Lee et al. (2023), we evaluate performance on three protein receptors: **fa7** (Coagulation factor VII), **5ht1b** (5-hydroxytryptamine receptor 1B), and **jak2** (Tyrosine-protein kinase JAK2). Ligand candidates are sampled from the learned distribution over the ZINC250k dataset (Irwin et al., 2012). To assess binding affinity, we compute docking scores using the program QuickVina 2 (Alhossary et al., 2015) and set the exhaustiveness to 1 by following Lee et al. (2023).

Evaluation metrics We assess ligand-protein binding based on four criteria: (1) Tanimoto similarity (**SIM**) with the closest ZINC250k training sample is below 0.4 to ensure novelty (**Nov.**); (2) synthetic accessibility score (**SA**) is below 5, indicating ease of synthesis; (3) drug-likeness score (**QED**) exceeds 0.5; and (4) docking score (**DS**) is lower than the median DS of known actives. For consistent comparison and training, each raw score is normalized to lie within the range, $0 \leq \hat{\mathbf{SA}}, \hat{\mathbf{DS}} \leq 1$:

$$\hat{\mathbf{SA}} = \frac{10 - \mathbf{SA}}{9} \qquad \hat{\mathbf{DS}} = \frac{\mathbf{DS}}{\min(\mathbf{DS}_{train}) - 0.2} \tag{26}$$

Where higher values indicate better performance. The overall evaluation metric, **Novel Hit Ratio (%)**, is the percentage of unique (**Uniq.**) molecules among 3,000 samples that satisfy these criteria.

Model and baselines We adopt the pretrained score-based generative model GDSS (Jo et al., 2022) and its classifier-guided variant MOOD (Lee et al., 2023) as baselines. We also compare SDG with several non-diffusion baselines: FREED (Yang et al., 2021), a fragment-based reinforcement

learning method; HierVAE (Jin et al., 2020), a VAE model with a hierarchical molecular representation; and MORLD (Jeon & Kim, 2020), a reinforcement learning approach that incorporates QED, SA, and DS at different optimisation stages. We also report ablations of SDG: *SDG w/o Stein correction*, corresponding to standard training-free diffusion guidance; *SDG (*$\alpha(t) = 0$*,* $\epsilon(t) > 0$*)*, corresponding to guidance without low-density sampling; and *SDG (*$\alpha(t) > 0$*,* $\epsilon(t) = 0$*)*, corresponding to the Langevin correction (Cororally 3.5).

Novel diffusion prior on the order of molecule graphs In graph generative modeling, the sampling process often utilizes the marginal graph order data distribution. However, for a certain type of molecular properties, the distribution of desired molecules over graph order is usually nonuniform; i.e., molecules can behave differently in molecular property space according to their node cardinality. Here, we propose a novel prior on the order of graphs that prioritizes sampling graph order, whose training molecules exhibit desired docking scores on a target protein receptor.

$$p^{\dagger}(N_i) = \frac{|N_i \cap (\mathbf{DS}_i < \tau)| + M \times 39}{M \times 39 + \sum_i N_i} \times \frac{p(N_i)}{p(\mathbf{DS} < \tau)} \tag{27}$$

Where $|\cdot|$ denotes the cardinality of set satisfying the N_i number of nodes and their docking scores $\mathbf{DS}_i$ below the hit threshold τ, i.e, $|N_i \cap (\mathbf{DS}_i < \tau)|$; M denotes the offset number that serves to enable sampling the graph order which does not have any molecules within desired docking scores; we choose $M = 10$ for all experiments; in ZINC250k dataset, the range of graph order is from 0 to 38, which results to 39 different possibilities; $p(N_i)$ is the marginal graph order distribution computed from training data; and $p(\mathbf{DS} < \tau)$ denotes the marginal distribution of hit training molecules.

Property predictor pretraining on clean data Since there are no available predictors for multiple target properties, we opt to pre-train our molecular property predictors on clean (i.e., noise-free) molecular data of ZINC250k, where the target property is defined as the product of the normalized scores: $\hat{\mathbf{SA}} \times \hat{\mathbf{DS}}$. Since most training molecules satisfy the **QED** hit condition, we thus ignore this target to simplify our multi-objective optimization task. Our regressor architecture is similar to the one from Lee et al. (2023) with an additional graph convolution layer. We set the learning rate to 0.01, the number of epochs to 10, the AdamW optimizer (Loshchilov & Hutter, 2019), and utilize the same architecture hyperparameters for all target proteins.

Stein correction stepsize We utilize an adaptive stepsize schedule $\epsilon(t)$ similar to the corrector framework from Song et al. (2020b), which is defined as:

$$\epsilon(t) = 2\eta^2(t) \left(snr\|\mathbf{z}\|_2 / \|\mathbf{g}\|_2\right)^2, \qquad \mathbf{z} \sim \mathcal{N}(0, I) \tag{28}$$

where snr denotes the signal-to-noise ratio, and $\mathbf{g} = \mathbf{s}_\theta\left(\mathbf{x}_T^j\right) - \eta(t)\mathbf{s}_\theta\left(\mathbf{x}_t^j\right)$, assuming a forward noising kernel as $p_{t|T}(\mathbf{x}_t|\mathbf{x}_T) = \mathcal{N}(\eta(t)\mathbf{x}_T, \gamma^2(t)I)$.

Annealing schedules For low-density sampling, we experiment on two scheduling approaches: *constant* and *linear*.

$$\alpha(t) = \begin{cases} \alpha_{max}, & \textit{constant} \\ t \times \alpha_{max}, & \textit{linear} \end{cases} \tag{29}$$

For guidance strength, we adopt the Polyak stepsize (Hazan & Kakade, 2019; Shen et al., 2024):

$$\beta(t) = \beta_{max} \times \frac{\|s_\theta(\mathbf{x}_t)\|}{\|\nabla_{\mathbf{x}_t} r(\mathbf{x}_T)\|} \tag{30}$$

Table 4: SDG hyperparameters.

	fa7	jak2	5ht1b
$snr_{\mathbf{X}}$	0.2	0.35	0.3
$snr_{\mathbf{A}}$	0.2	0.35	0.3
α scheduling	*constant*	*constant*	*constant*
$\alpha_{\mathbf{X}_{max}}$	0.20	0.42	0.35
$\alpha_{\mathbf{A}_{max}}$	0.20	0.42	0.35
$\beta_{\mathbf{X}_{max}}$	1.	1.	.7
$\beta_{\mathbf{A}_{max}}$	0	0	0
N	512	3000	512

Table 5: Mean and standard deviation of novel top 5% docking scores over three sampling runs. Baseline results are taken from Lee et al. (2023).

Method	Top 5% DS (↓)		
	fa7	5ht1b	jak2
HierVAE Jin et al. (2020)	−6.812 (±0.274)	−8.081 (± 0.252)	−8.285 (±0.370)
MORLD Jeon & Kim (2020)	−6.263 (±0.165)	−7.869 (± 0.650)	−7.816 (±0.133)
FREED Yang et al. (2021)	−8.297 (±0.094)	−10.425 (± .331)	−9.624 (±0.102)
GDSS Jo et al. (2022)	−7.775 (±0.039)	−9.459 (± 0.101)	−8.926 (±0.089)
MOOD Lee et al. (2023)	−8.160 (±0.071)	−11.145 (±0.042)	−10.147 (±0.060)
SDG w/o Stein correction	−7.794 (±0.040)	−7.370 (±0.201)	−5.904 (±0.163)
SDG	−8.310 (±0.009)	−11.383 (±0.0537)	−10.178 (±0.037)

Hyperparameter search We conducted a hyperparameter search on SDG. We set the signal noise ratio $snr = \{0.2, 0.3, 0.35\}$, the maximum low-density level $\alpha = \{0.1, 0.2, 0.35, 0.42\}$, the maximum guidance strength $\beta = \{0.5, 0.7, 1.0\}$, the number of particles $N = \{512, 1024, 3000\}$, and the anneal scheduling $\alpha_{\text{scheduling}} = \{\text{linear}, \text{constant}\}$. Table 4 presents the hyperparameters of the main experimental results.

Hardware usage All experiments are conducted on NVIDIA Titan RTX, RTX 6000 with 8 CPU cores, using a Slurm-managed high-performance computing (HPC) system.

C.3 Additional results

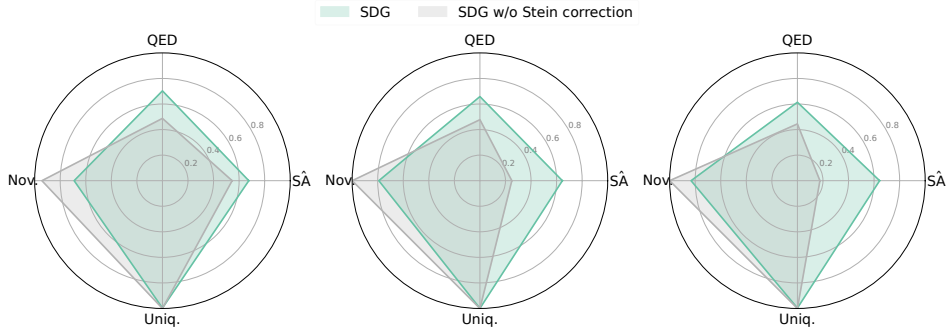

Figure 8: SDG performance on radar plots: fa7 (left), jak2 (middle), 5ht1b (right).

C.3.1 MULTI-OBJECTIVE OPTIMIZATION

Figure 8 displays the radar plots of SDG performance under multiple property constraints. A consistent pattern emerges across all target proteins: without Stein correction, SDG tends to generate novel molecules with low drug-likeness and poor synthetic, largely due to overly complex structures. Notably, SDG also achieves nearly 100% molecular uniqueness.

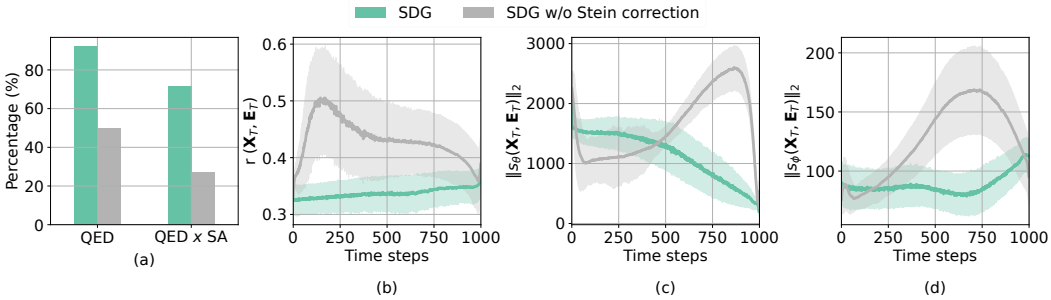

Figure 9: Temporal sampling dynamics of SDG for the fa7 protein. (a) Percentage of molecules meeting QED and SA hit criteria; (b) Rewards of posterior samples $r(\mathbf{X}_T, \mathbf{E}_T)$; (c) Frobenius norm of node posterior scores $s_\theta(\mathbf{X}_T, \mathbf{E}_T)$; (d) Frobenius norm of edge posterior scores $s_\phi(\mathbf{X}_T, \mathbf{E}_T)$.

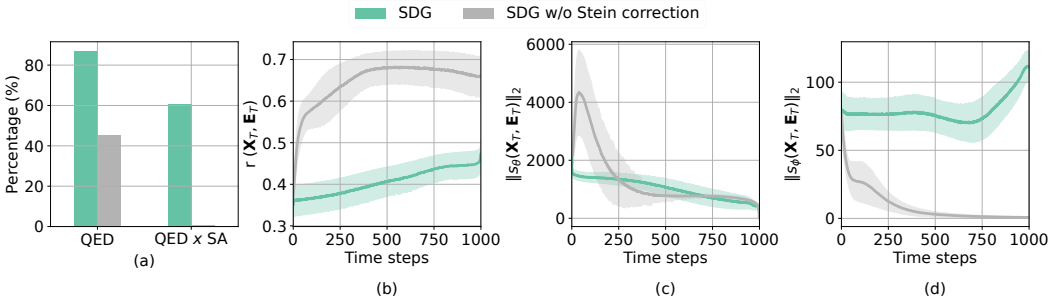

Figure 10: Temporal sampling dynamics of SDG for the jak2 protein. (a) Percentage of molecules meeting QED and SA hit criteria; (b) Rewards of posterior samples $r(\mathbf{X}_T, \mathbf{E}_T)$; (c) Frobenius norm of node posterior scores $s_\theta(\mathbf{X}_T, \mathbf{E}_T)$; (d) Frobenius norm of edge posterior scores $s_\phi(\mathbf{X}_T, \mathbf{E}_T)$.

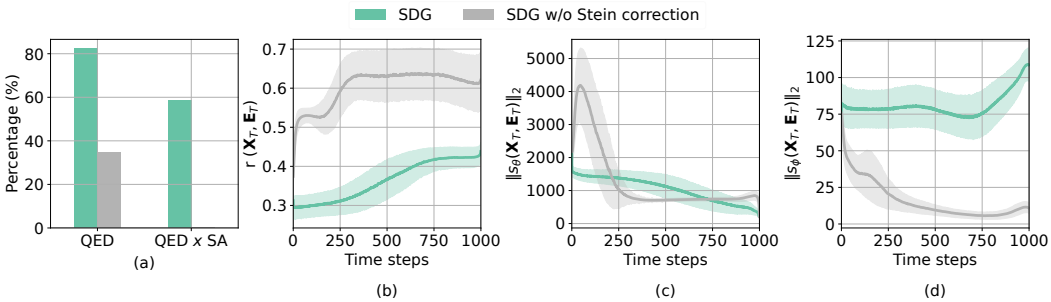

Figure 11: Temporal sampling dynamics of SDG for the 5ht1b protein. (a) Percentage of molecules meeting QED and SA hit criteria; (b) Rewards of posterior samples $r(\mathbf{X}_T, \mathbf{E}_T)$; (c) Frobenius norm of node posterior scores $s_\theta(\mathbf{X}_T, \mathbf{E}_T)$; (d) Frobenius norm of edge posterior scores $s_\phi(\mathbf{X}_T, \mathbf{E}_T)$.

C.3.2 NOVEL TOP 5% DOCKING SCORES

We additionally report in Table 5 the average docking scores of the top 5% unique molecules that satisfy the novel hit conditions. As observed, SDG yields significantly lower average docking scores compared to the standard training-free diffusion guidance method (SDG w/o Stein correction), indicating that the generated molecules exhibit a stronger binding affinity. Moreover, SDG improves binding affinity over both the pretrained model GDSS and the classifier-based diffusion guidance method MOOD.

C.3.3 SAMPLING DYNAMICS OF MODEL SCORES ON $\mathcal{M}_T$

In the experiments, we only apply the guidance on the node's score model, $\beta_X(t) > 0$, and ignore the guidance on the edge's score model, $\beta_A(t) = 0$ for all t. This enables low-cost sampling by avoiding backpropagation through edge/adjacency matrices, whose dimension is $\mathcal{O}(N^2)$, with N being the number of nodes. Notably, thanks to the system of coupled SDEs (Equation 23), any updates on the node component will be appropriately reflected on the edge component as well; controlling the node-component reward guidance would thus be expressive enough to sample molecular graphs with desired properties.

During sampling process, the model scores $s(\cdot)$ tend to *increase* in norm as samples move toward the data manifold, i.e, $\|s(\mathbf{x_{t+1}})\|_2 > \|s(\mathbf{x_t})\|_2$, with $x_{t+1} \in \mathcal{M}_{t+1} \cup x_t \in \mathcal{M}_t$. In contrast, the model scores on posterior samples tend to *decrease*, i.e, $\|s(\mathbf{x}_T|\mathbf{x_{t+1}})\|_2 < \|s(\mathbf{x}_T|\mathbf{x_t})\|_2$, under the same manifold transition, reflecting that posterior samples move closer to the data distributions. These can be intuitively observed via a single data point case $\mathbf{x}^\dagger$ with Gaussian noises, where the normed conditional model score $\|s(\mathbf{x}, t)\|_2 \propto \left\| \frac{\mathbf{x}-\mathbf{x}^\dagger}{\sigma_t^2} \right\|_2$.

$$\begin{aligned} &\text{Toward the data manifold:} \quad t \to T, \quad \sigma_t \to \sigma_{min}, \quad \mathbf{x} \quad \Rightarrow \quad \|s(\mathbf{x}, t)\|_2 \nearrow \\ &\text{On the data manifold:} \quad t = T, \quad \sigma_{min}, \quad \mathbf{x} \to \mathbf{x}^\dagger \quad \Rightarrow \quad \|s(\mathbf{x}, t)\|_2 \searrow \end{aligned} \tag{31}$$

Figures 9, 10, and 11(c,d) illustrate the sampling dynamics of model scores. Without the Stein correction, score dynamics fluctuate arbitrarily across both edge and node components. In contrast, with the Stein correction, node scores smoothly transit toward high-density regions, while edge scores first converge to high-density regions within the initial 700–750 steps before shifting toward lower-density regions in the opposite direction. These dynamics align well with the theoretical analysis of diffusion score behavior, providing further evidence that the Stein correction regularizes diffusion guidance in low-density regions.

C.3.4 REWARD OVERESTIMATION IN DIFFUSION GUIDANCE

In training-free diffusion guidance, reward models and classifiers are trained only on a finite set of clean data samples, making their reward estimations reliable within the data support. However, diffusion models have much broader support due to noise injection, which can push samples outside the data support during sampling. In such regions, reward models often overestimate rewards, producing unreliable guidance. Uehara et al. (2024) were the first to formalize this issue, showing that reward models can assign excessively high scores to samples whose semantics or properties fail to meet the desired criteria. To mitigate this, they proposed regularizing the reverse diffusion process via the original stochastic optimal control formulation—though this approach is computationally expensive and impractical for efficient sampling.

Motivated by this challenge, Stein Diffusion Guidance provides a more practical alternative by solving the surrogate stochastic optimal control objective. As shown in Figures 9, 10, and 11 (a, b), the absence of Stein correction leads the reward models to overestimate genuine rewards, producing artificially inflated values. However, the corresponding molecules are often unrealistic, with significantly lower QED and SA scores. By contrast, the Stein correction introduces a low-cost regularization that mitigates reward overestimation. Furthermore, SDG-regularized model scores evolve smoothly throughout the sampling process (see Section C.3.3), keeping sampling trajectories within generative manifolds and enabling effective exploration in low-density regions.

C.3.5 ABLATION STUDY ACROSS OODNESS LEVELS

We evaluate the robustness of SDG under varying levels of out-of-distribution (OOD) sampling, where the maximum low-density level ($\alpha_{\max}$) defines the OODness level. To compare generated and test samples, we use distribution-based metrics in chemical space (**FCD**) and structural space (**NSPDK**) (Lee et al., 2023). Figures 12, 13, and 14 (a,b) illustrate the results for SDG and standard training-free guidance (SDG w/o Stein correction). As OODness increases, standard training-free guidance drifts completely off the data manifold in both chemical and structural spaces, yielding significantly high FCD and NSPDK values. In contrast, SDG remains closer to the data manifold, with only marginal increases in distribution distances, thereby enabling effective sampling in low-density regions. Notably, even under an extreme OODness level ($\alpha = 0.5$), SDG preserves a high rate of valid valency (**Validity**) in the generated molecules, as shown in Figures 12, 13, and 14 (c).

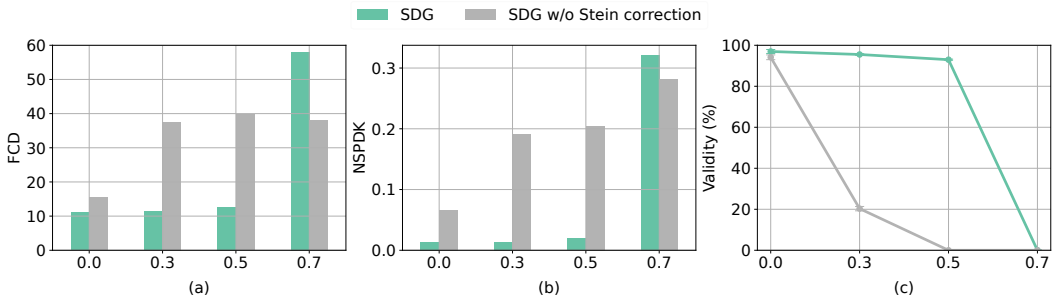

Figure 12: Ablation results under different low-density levels ($\alpha_{\max}$) for fa7. Chemical distance FCD (a) and structural distance NSPDK (b) to the test set; (c) Validity of generated molecules.

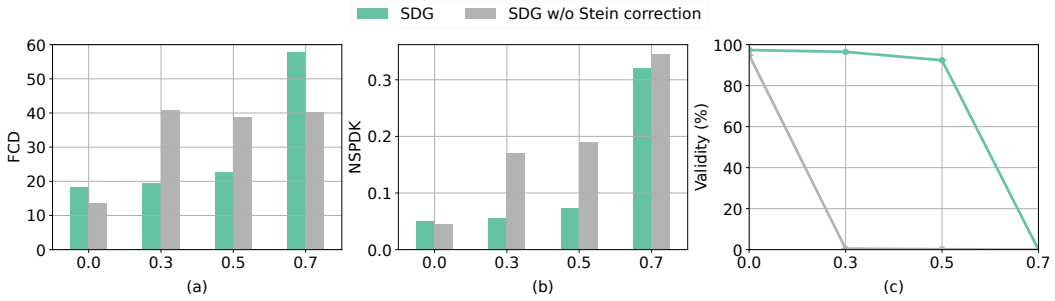

Figure 13: Ablation results under different low-density levels ($\alpha_{\max}$) for jak2. Chemical distance FCD (a) and structural distance NSPDK (b) to the test set; (c) Validity of generated molecules.

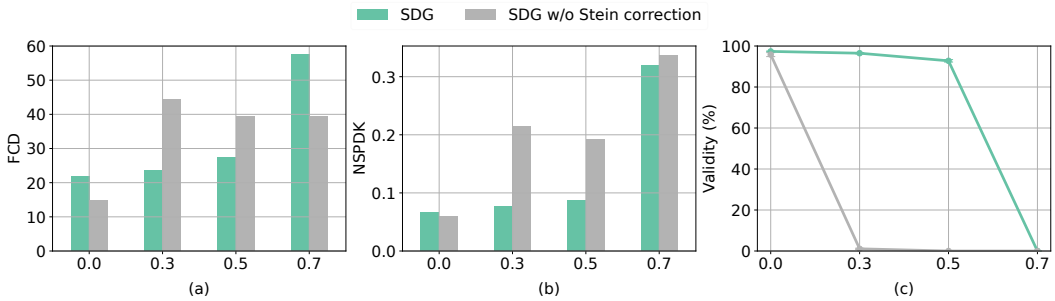

Figure 14: Ablation results under different low-density levels ($\alpha_{\max}$) for 5ht1b. Chemical distance FCD (a) and structural distance NSPDK (b) to the test set; (c) Validity of generated molecules.

C.3.6 COMPUTATIONAL ANALYSIS

SDG provides a low-cost alternative to the stochastic optimal control framework of Uehara et al. (2024) for diffusion guidance sampling. Whereas the prior approach incurs a worst-case runtime complexity of $\mathcal{O}(T^2)$ due to requiring full trajectory simulations at each step, SDG achieves linear complexity $\mathcal{O}(T)$, with only a modest overhead relative to standard training-free guidance. We benchmarked on an RTX 6000 with 8 CPU cores to sample 3000 molecules, using the same solver as the pretrained score-based models (Jo et al., 2022; Lee et al., 2023). Table 6 reports SDG's computational cost in both settings. The runtime roughly doubles with the inclusion of the Stein correction, though this overhead can be mitigated with faster solvers such as DDIMs (Song et al., 2020a). Importantly, the back-and-forth Stein correction reduces memory overhead, yielding more efficient usage compared to the original control problem (Uehara et al., 2024).

Table 6: Computation analysis on the molecule guidance task.

	RUNTIME (SECOND)	MEMORY (MB)
SDG w/o Stein correction	1360	7783
SDG ($\alpha(t) > 0, \epsilon(t) > 0$)	2521	8049

D IMAGE DIFFUSION GUIDANCE TASKS

We evaluate the generalizability of SDG to diffusion guidance tasks beyond molecules, particularly on images. These experiments highlight SDG's potential applicability across diverse data domains. Following the unified training-free diffusion guidance framework (TFG) of Ye et al. (2024), we test SDG on three representative tasks: label guidance, Gaussian deblurring, and image super-resolution:

- *Label guidance:* sampling images with desired class labels from the CIFAR-10 dataset.
- *Gaussian deblurring:* reconstructing original images that have been blurred with Gaussian kernels.
- *Image super-resolution:* reconstructing high-resolution images at 256×256 from their downsampled 64×64 counterparts.

Evaluation metrics We evaluate training-free diffusion guidance methods using standard image metrics. **Accuracy** (%) measures the average classification accuracy on generated samples across CIFAR-10 labels, while **FID** and **LPIPS** (Zhang et al., 2018) assess the fidelity and perceptual similarity between generated and test images.

Model setup and baselines To adapt TFG from Ye et al. (2024), we remove the recurrent strategy (line 5) and mean guidance (line 8) from their Algorithm 1; the remaining components correspond to a standard training-free guidance approach incorporating the implicit dynamic control of LGD (Song et al., 2023) (line 4). We retain the guidance hyperparameters from their original parameter search. Since these tasks involve no minority image classes, we set the low-density guidance factor to zero ($\alpha = 0$) and focus solely on optimizing samples with highly desired properties. We also compare our results with DPS (Chung et al., 2023), a baseline closely related to SDG without the Stein correction, but lacking the implicit dynamic control.

Computation analysis We report computation details for the label guidance task. Experiments were conducted on an RTX 3090 GPU with 4 CPU cores and a batch size of 256. Table 7 summarizes the runtime and memory usage for sampling 2,560 images per class over $T = 100$ DDIM steps (Song et al., 2020a). The back-and-forth Stein correction notably reduces memory overhead. Moreover, the runtime overhead decreases compared to the previous solvers in Table 6, from $185\% = \frac{2521}{1360} \times 100$ to $136\% = \frac{820}{602} \times 100$.

Table 7: Computation analysis on the image label guidance task.

	RUNTIME (SECOND)	MEMORY (MB)
SDG w/o Stein correction	602	18728
SDG ($\alpha(t) > 0, \epsilon(t) > 0$)	820	18792

E VISUALIZATION

E.0.1 VISUALIZATION OF IMAGE DIFFUSION GUIDANCE TASKS

Figure 15: Visualization of image deblurring results: SDG without Stein correction (Left) vs. SDG with Stein correction (Right).

Figure 16: Visualization of image super-resolution results: SDG without Stein correction (Left) vs. SDG with Stein correction (Right).

E.0.2 VISUALIZATION OF LIGAND-PROTEIN DOCKING POSES

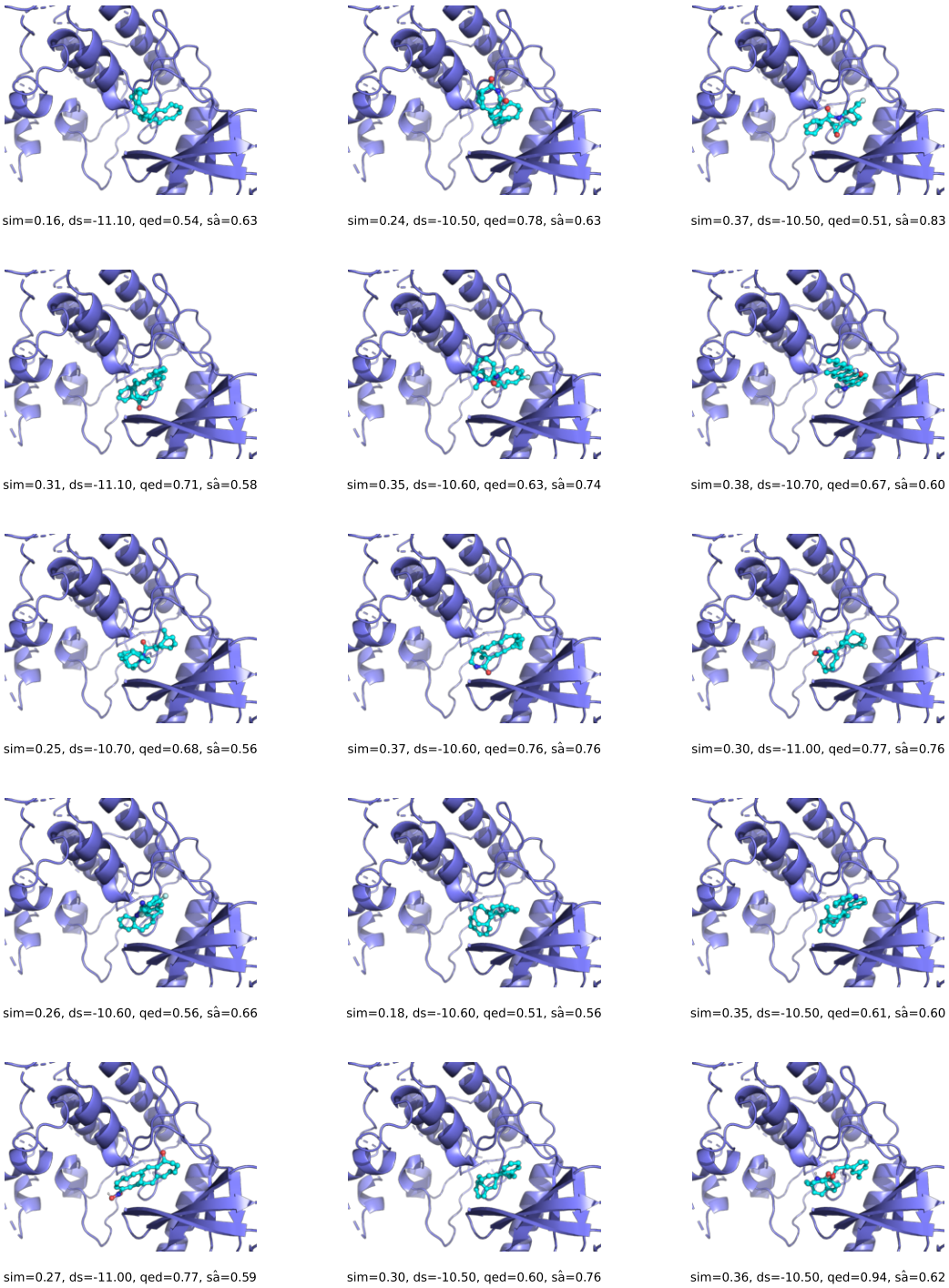

Figure 17: Visualization of docking poses for multiple generated ligands bound to the jak2 protein.

Figure 18: Visualization of docking poses for multiple generated ligands bound to the fa7 protein.

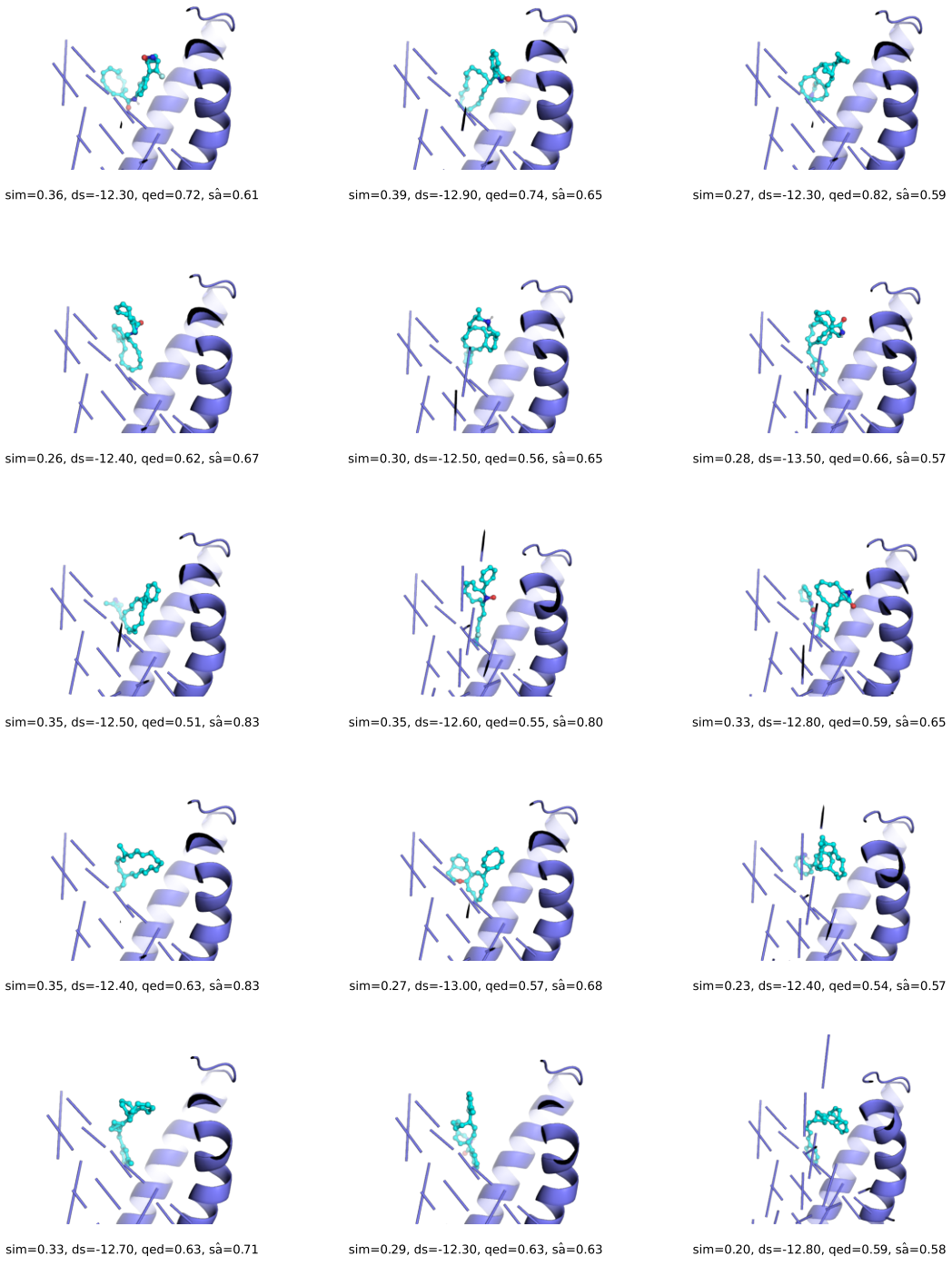

Figure 19: Visualization of docking poses for multiple generated ligands bound to the 5ht1b protein.

F Declaration of LLM usages

We only leverage LLM tools for editing purposes, which mainly include grammar checks, spelling corrections, and word choice suggestions.