Briefings in Bioinformatics, 2026, 27, bbag033
https://doi.org/10.1093/bib/bbag033
Published: 5 February 2026
Problem Solving Protocol

OXFORD

Global and local integrated gradient-based diffusion model for *de novo* drug design

Sejin Park[1], Minjae Chung[2], Hyunju Lee[1,2,3,*]

[1]Department of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu,Gwangju 61005, Republic of Korea
[2]Department of Artificial Intelligence Convergence, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu,Gwangju 61005, Republic of Korea
[3]GIST InnoCORE AI-Nano Convergence Institute for Early Detection of Neurodegenerative Diseases, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Republic of Korea

*Corresponding author. E-mail: hyunjulee@gist.ac.kr

Abstract

In *de novo* drug design, deep learning-based approaches have become essential to efficiently navigate the vast chemical space of drug-like molecules. Recently, diffusion-based models have attracted significant attention in the generation of target-binding molecules. However, these models have difficulty in simultaneously optimizing the binding affinity and drug-like properties and require high computational costs because of the long and sequential denoising process. To address these limitations, we propose the Global and local integrated gradient-based Diffusion Model (GlintDM). GlintDM introduces a significantly faster denoising process, namely *skip transition*, by leveraging global gradients and local gradients. Due to the fast denoising process, GlintDM can perform the following three phases during the molecule generation: position refinement, candidate evaluation, and ligand resampling. These phases allow GlintDM to identify optimal binding positions to the target protein and generate molecules satisfying multi-objective molecular properties. As a result, GlintDM outperforms other methods on both the CrossDocked and Binding MOAD datasets for Vina-related scores. Further validation through the PoseBusters test and assessment of molecular properties, such as steric clash and geometric properties, confirm that GlintDM can generate stable and high-quality molecules.

Keywords diffusion model, drug discovery, generative model, structure-based drug design

Introduction

In *de novo* drug discovery, artificial intelligence (AI)-based methods have become powerful tools. For instance, they enable the rapid exploration of vast chemical spaces, an otherwise infeasible task using conventional approaches, thus facilitating the efficient identification of novel scaffolds with desirable pharmacological properties [1, 2].

In *de novo* drug design, molecular representations are commonly categorized into three main approaches: string-based, 2D graph-based, and 3D structure-based methodologies. The string-based approach [3, 4] represents molecules using SMILES [5] and typically employs language models, such as recurrent neural networks [6]. While this method enables efficient data handling and large-scale dataset construction due to its simplicity, it lacks the ability to encode detailed geometric and spatial information crucial for molecular interactions. In contrast, 2D graph-based methods [7, 8] represent atoms as nodes and bonds as edges, allowing the use of graph neural networks for molecular modeling. These models can capture structural and relational information that extends beyond the capabilities of SMILES, yet they still lack the ability to incorporate explicit 3D features. Recently, in 3D structure-based drug design (SBDD), diffusion-based generative models have demonstrated strong performance and have been explored through a variety of methodological approaches [9–11].

Diffusion models can be broadly categorized into denoising diffusion probabilistic models (DDPMs) [12] and score matching with Langevin dynamics (SMLD) [13]. DDPM generates new samples by progressively removing noise from initial noisy samples. In contrast, SMLD assumes that clean data points reside in high-probability regions of the data distribution, whereas noisy data points are found in low probability areas. Using score matching to estimate the gradient of the data distribution, generative models based on SMLD can move the point of the noisy data toward the high-probability density regions and generate clean samples. Although some molecular generative models [14, 15] adopt SMLD, the majority [9, 10, 16, 17] rely on DDPM

Received: September 27, 2025. **Revised:** November 25, 2025. **Accepted:** January 9, 2026

due to the difficulty in directly defining gradients in the molecular data distribution.

Despite differences in methodology, both DDPM and SMLD aim to iteratively transit data points from $T = t$ to $t - 1$, resulting in substantial computational overhead and inherently slow generation. To alleviate the slow diffusion process, denoising diffusion implicit models (DDIMs) [18] and consistency models (CMs) [19] have been introduced. DDIMs accelerate the sampling process of diffusion models by removing the Markov assumption and introducing a deterministic generative path, and CMs leverage consistency regularization to enable one-step or a few-step sampling. Although these approaches significantly reduce the number of diffusion transition steps, DDIMs still require multiple steps, and CMs may suffer from maintaining distribution accuracy.

Most diffusion-based molecular generative models utilize 3D molecular structures to represent molecules, employing SE(3)- or E(3)-equivariant neural networks such as EGNNs and SE(3)-Transformers [20, 21] to process these geometrically structured data. Early diffusion-based generative models include EDM [22] for property-conditioned generation, TargetDiff [9] for target-specific molecule design, and DiffHopp [16] for scaffold hopping. Building upon DiffHopp, TurboHopp [11] integrates CMs [19] to accelerate the diffusion process and leverages reinforcement learning to improve target-related objective scores. However, it is restricted to scaffold hopping and requires additional training to generate molecules satisfying multi-objective properties.

Inspired by TargetDiff, several approaches have been proposed to enhance target-binding molecule generation, including UniGuide [23], BADGER [24], TAGMOL [25], and IPDIFF [26]. Specifically, BADGER, IPDIFF, and TAGMOL employ guidance networks to steer the generation process toward protein-binding ligands. In contrast, UniGuide adopts a unified self-guidance approach based solely on geometric constraints, avoiding the need for additional networks to predict binding affinity. These models [24–26] have incorporated binding affinity information (e.g. gradient guidance or mean shift updates) into the latent distribution to approach the proper optimal point. However, this information was only reflected in the atom coordination distribution, not the atom type distribution.

Diffusion-based models have demonstrated remarkable performance in generating ligands with higher binding affinity scores compared with string- or 2D graph-based generative models. Despite these advances, they exhibit two inherent limitations: (i) substantial computational cost and (ii) difficulty in simultaneously optimizing multiple objectives (e.g. high binding affinity and drug-likeness). Fundamentally, the diffusion process requires a large number of incremental steps, rendering it computationally intensive and inherently constrained in its ability to explore the chemical space globally. Although refinements and resampling techniques have improved stability and coherence in inpainting and generative tasks, particularly for ligand–protein complexes [10, 23, 27–31], the necessity of numerous diffusion transitions remains a critical bottleneck.

To overcome these drawbacks, we propose a **GlintDM**—**G**lobal and **l**ocal **int**egrated gradient-based **D**iffusion **M**odel. GlintDM leverages both global and local gradients of atom positions and types to accelerate the denoising process, a strategy we refer to as *skip transition*. The introduction of *skip transitions* significantly reduces the number of denoising steps, thereby enabling efficient application of multiple refinements and resampling. In detail, GlintDM follows the backbone of TargetDiff [9], which is built upon E(3)-equivariant graph neural networks [20], and consists of three distinct phases: *position refinement* to identify optimal binding sites, *candidate evaluation* to discover promising regions of the data distribution, and *ligand resampling* to generate stable ligands with desirable molecular properties (Fig. 1). These three phases allow GlintDM to generate stable and multi-objective molecules for target proteins.

On the CrossDocked and Binding MOAD datasets, GlintDM outperformed recently developed methods in Vina-related scoring. PoseBusters evaluations and analyses of various molecular properties further confirmed that GlintDM can generate stable and high-quality multi-objective molecules for target proteins. Moreover, on the Binding MOAD dataset, GlintDM achieved significantly faster generation speeds—approximately four times faster than TargetDiff and eight times faster than TAGMOL.

Methods

Notations and problem definition

A protein and a ligand are represented as the graph $P = \{\mathbf{x}^p, \mathbf{h}^p\}$ and $L = \{\mathbf{x}, \mathbf{h}\}$, respectively, where $\mathbf{x} \in \mathbb{R}^{N\times 3}$ and $\mathbf{h} \in \mathbb{R}^{N\times F}$ are the Cartesian coordinates and F-dimensional feature vectors, respectively. To be specific, the feature vector of proteins consists of the atom elements (6), amino acid types (20), and backbone indicators (1), and the ligand feature vector is the one-hot vector of atom types with aromaticity (13). In this work, we treat the protein as having a fixed conformation and focus only on protein pocket sites, where atoms are within 10 Å region around the reference ligand. Our goal is to generate ligands binding to the target protein pocket. In addition, we translate the complex so that the center of mass (CoM) of the protein atoms is located at the origin. This places the system in a CoM-free coordinate frame. This normalization step is essential for establishing an SE(3)-invariant initial density, ensuring that the diffusion process is unaffected by global translations and depends only on the relative geometry of the complex.

Molecular diffusion process

Given samples from a data distribution $q(L_0|P)$, a diffusion model is trained to approximate the model distribution $p_\theta(L_0|P)$ to match the data distribution $q(L_0|P)$. Following DDPM [12], the form of the latent variable model is

$$p_\theta(L_0|P) = \int p_\theta(L_{0:T}|P)\mathbf{d}L_{1:T}, \tag{1}$$

where $L_1, L_2, ..., L_T$ are latent variables that share the same sample space as the data $L_0 \sim q(L_0|P)$. The diffusion model learns the forward diffusion process, $q(L_{1:T}|L_0, P)$, and reverse generative process, $p_\theta(L_{0:T-1}|L_T, P)$. These processes are defined as Markov chains:

$$\begin{aligned} q(L_{1:T}|L_0, P) &= \prod_{t=1}^{T} q(L_t|L_{t-1}, P), \\ p_\theta(L_{0:T-1}|L_T, P) &= \prod_{t=1}^{T} p_\theta(L_{t-1}|L_t, P). \end{aligned} \tag{2}$$

To sample the ligand graph $L = \{\mathbf{x}, \mathbf{h}\}$, we use Gaussian distributions, $\mathcal{N}$, and the Gumbel-Softmax trick [32], $q(\mathbf{h})$, to model continuous atom coordinates $\mathbf{x}$ and atom type probabilities $\mathbf{h}$, respectively. Here,

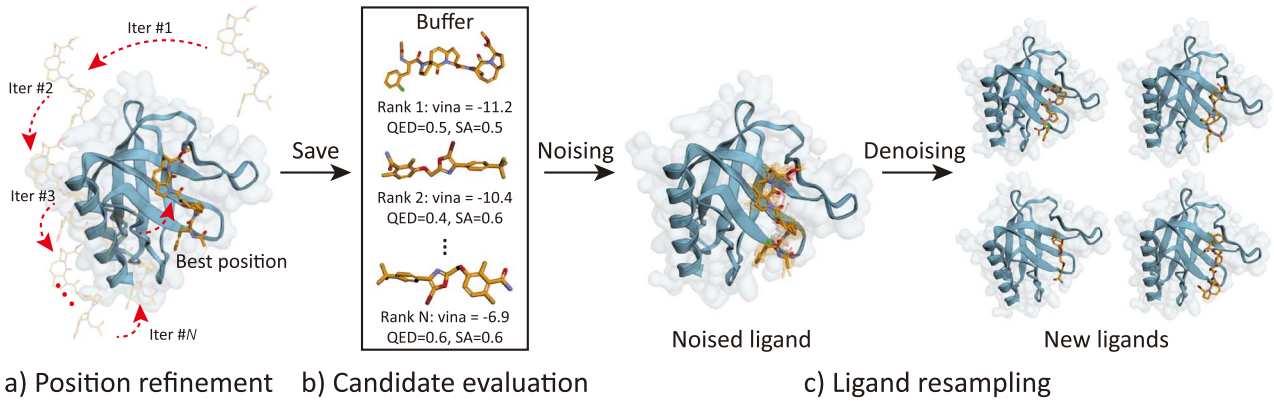

Figure 1 Overview of GlintDM, illustrating (a) Position refinement: the multiple diffusion procedures are iteratively applied to identify the most probable binding site, (b) Candidate evaluation: promising ligand candidates are evaluated and stored in the buffer, (c) Ligand resampling: noisy candidate ligands are used as input to the diffusion model, and then new ligands are generated through the denoising process.

we define the ligand distribution as a product of $\mathcal{N}$ and $q(\mathbf{h})$:

$$q(\mathbf{h}_t|\mathbf{h}_{t-1}) := \text{LogSoftMax}(\log((1-\beta_t)\cdot\exp(\mathbf{h}_{t-1}) + \beta_t \cdot \exp(\mathbf{g}))) \quad (3)$$

$$q(L_t|L_{t-1}, P) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \cdot q(\mathbf{h}_t|\mathbf{h}_{t-1}), \quad (4)$$

where $\beta_1, ..., \beta_T$ are fixed variance schedules, and $\mathbf{g} \sim$ Gumbel(0, C). In the Gumbel distribution, the constant C controls the scale of the noise; i.e. a larger C increases the likelihood of producing larger noise values, whereas a smaller C reduces this likelihood. Although atom positions and types are inherently dependent, we assume their distributions to be independent for mathematical simplicity. Under this assumption, we efficiently train the neural network to learn both forward and reverse processes. However, in practice, the two distributions influence each other within the same network, effectively capturing their underlying dependencies.

The noisy data distribution $q(L_t|L_0)$ of any time step in closed-form can be calculated by the property of the diffusion process:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}),$$

$$q(\mathbf{h}_t|\mathbf{h}_0) := \text{LogSoftMax}\left(\log\left(\bar{\alpha}_t \cdot \exp(\mathbf{h}_0) + (1-\bar{\alpha}_t)\cdot\exp(\mathbf{g})\right)\right), \quad (5)$$

where $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$, and $\mathbf{g} \sim$ Gumbel(0, 1).

Using Bayes' theorem, the normal posterior of atom coordinates and atom type probabilities can both be computed in closed-form:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t\mathbf{I}), \quad (6)$$

where $\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t$ and $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$.

$$q(\mathbf{h}_{t-1}|\mathbf{h}_t, \mathbf{h}_0) = \frac{q(\mathbf{h}_t|\mathbf{h}_{t-1}, \mathbf{h}_0)q(\mathbf{h}_{t-1}|\mathbf{h}_0)}{\sum_{\mathbf{h}_{t-1}} q(\mathbf{h}_t|\mathbf{h}_{t-1}, \mathbf{h}_0)q(\mathbf{h}_{t-1}|\mathbf{h}_0)}$$

$$= \frac{q(\mathbf{h}_t|\mathbf{h}_{t-1})q(\mathbf{h}_{t-1}|\mathbf{h}_0)}{\sum_{\mathbf{h}_{t-1}} q(\mathbf{h}_t|\mathbf{h}_{t-1})q(\mathbf{h}_{t-1}|\mathbf{h}_0)} =: \tilde{c}_{t-1}(\mathbf{h}_t, \mathbf{h}_0), \quad (7)$$

where $q(\mathbf{h}_{t-1}|\mathbf{h}_0) := \text{LogSoftMax}(\log(\bar{\alpha}_{t-1}\cdot\exp(\mathbf{h}_0) + (1-\bar{\alpha}_{t-1})\cdot\exp(\mathbf{g})))$.

Algorithm 1 Training procedure of GlintDM

Schematic overview. This algorithm trains GlintDM by teaching the model to reconstruct the original molecular structure from noisy samples. At each step, noisy ligand coordinates ($\mathbf{x}_t$) and type logits ($\mathbf{h}_t$) are generated, and the network predicts the denoised estimates ($\hat{\mathbf{x}}_0$, $\hat{\mathbf{h}}_0$). Finally, the model parameters are updated using a combination of coordinate-reconstruction, KL divergence, and atom-type prediction losses.

Require: Protein–ligand dataset $\{\mathcal{P}, \mathcal{L}\}$, neural network ϕ_θ.

for ep in Epoch **do**
  Sample diffusion time $t \in \mathcal{U}(0, ..., T)$
  Move the complex to make CoM of protein atoms zero
  Diffuse coordinates: $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + (1-\bar{\alpha}_t)\epsilon$
  Diffuse atom types via Gumbel distribution
  Predict $[\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0] = \phi_\theta([\mathbf{x}_t, \mathbf{h}_t], t, P)$
  Compute posteriors for $\hat{\mathbf{h}}_{t-1}$ and $\mathbf{h}_{t-1}$ according to Equation 6.
  Compute loss: coordinate MSE + KL + NLL
  Update θ
end for

Training objective

The forward noising process is implemented by $\mathbf{x}_t = \sqrt{\alpha_t}\mathbf{x}_{t-1} + (1-\alpha_t)\epsilon$ and $\mathbf{h}_t = \text{LogSoftMax}(\log(\alpha_t\cdot\exp(\mathbf{h}_{t-1}) + (1-\alpha_t)\cdot\exp(\mathbf{g})))$. The reverse generative process is calculated by predicting $\hat{L}_0 = [\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0]$. Specifically, the neural network ϕ_θ (denoiser function) predicts $\hat{L}_0 = [\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0] = \phi_\theta([\mathbf{x}_t, \mathbf{h}_t], t, P)$, and then $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \hat{\mathbf{x}}_0, P)$ and $q(\mathbf{h}_{t-1}|\mathbf{h}_t, \hat{\mathbf{h}}_0, P)$ are calculated as the reverse process. In short, we can train the neural network by parameterizing $\mu_\theta(\mathbf{x}_t, P)$ and $c_\theta(\mathbf{h}_t, P)$, where three objective functions are used:

$$\mathcal{L}_{t-1}^{(\mathbf{h})} = \sum_k \tilde{c}_{t-1}(\mathbf{h}_t, \mathbf{h}_0)_k \log\frac{\tilde{c}_{t-1}(\mathbf{h}_t, \mathbf{h}_0)_k}{\tilde{c}_{t-1}(\mathbf{h}_t, \hat{\mathbf{h}}_0)_k} \quad (8)$$

$$\mathcal{L}_{NLL}^{(\mathbf{h})} = -\sum_i \mathbf{h}_0^i \log\hat{\mathbf{h}}_0^i \quad (9)$$

$$\mathcal{L}_{t-1}^{(\mathbf{x})} = \frac{1}{2\sigma_t^2}\|\tilde{\mu}_t(\mathbf{x}_t, \mathbf{x}_0) - \mu_\theta(\mathbf{x}_t, P)\|_2^2 + C$$

$$= \gamma_t\left\|\mathbf{x_0} - \hat{\mathbf{x}}_0\right\|_2^2 + C \approx \left\|\mathbf{x_0} - \hat{\mathbf{x}}_0\right\|_2^2, \quad (10)$$

where $\gamma_t = \frac{\bar{\alpha}_{t-1}\beta_t^2}{2\sigma_t^2(1-\bar{\alpha}_t)^2}$ and C is a constant.

Algorithm 2 Sampling procedure based on local and global gradients

Schematic overview. This algorithm samples ligand structures by combining global and local gradients:
(1) a *local gradient* guiding the transition to the next timestep
(2) a *global gradient* directing the estimated final state.
By integrating these gradients, GlintDM iteratively refines both coordinates and atom types while skipping several intermediate steps.

Require: Binding-site P, pretrained model ϕ_θ

1: Sample number of ligand atoms and initialize atom coordinates/-types
2: **for** t in $T, T-n, \ldots, 1$ **do**
3: Predicts denoised $[\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0] = \phi_\theta([\mathbf{x}_t, \mathbf{h}_t], t, P)$
4: Sample $\mathbf{h}_{t-1}$ and $\mathbf{x}_{t-1}$ from posteriors according to Equations 6 and 7.
5: Compute local gradients $(\nabla\mathbf{x}_l, \nabla\mathbf{h}_l)$
6: Compute global gradients $(\nabla\mathbf{x}_g, \nabla\mathbf{h}_g)$
7: Intergrate gradients using $\lambda_t = t/T$
8: Update coordinates and types for step $t-n$
9: **end for**

Equation 8 is used to learn the probability distribution of ligand atom types during the denoising step from $T = t$ to $t-1$, by aligning the estimated probabilities with the reference posterior distribution. Equation 9 captures the atom type probabilities throughout the denoising trajectory from $T = t$ to 0. In this case, by minimizing the negative log-likelihood (NLL) loss, the network ϕ is trained to directly predict atom types of the final state ($T = 0$) from the state ($T = t$). In summary, Equations 8 and 9, respectively. capture global and local transitions in the categorical atom type distribution. Since $\mathbf{x}_{t-1}$ can be derived from $\mathbf{x}_t$ and $\hat{\mathbf{x}}_0$, Equation 10 facilitates accurate reverse inference of the positional distribution. By minimizing the discrepancy between the predicted denoised position $\hat{\mathbf{x}}_0$ and the ground truth $\mathbf{x}_0$, it effectively captures both global and local transitions. Therefore, the final loss function is $\mathcal{L} = \mathcal{L}_{t-1}^{(\mathbf{h})} + \mathcal{L}_{NLL}^{(\mathbf{h})} + \mathcal{L}_{t-1}^{(\mathbf{x})}$, and the training details are provided in Algorithm 1 and S1.

Combining global and local gradients

Conventional diffusion models operate by iteratively transitioning data from a state at time t to $t-1$. A principal limitation of this methodology is the requirement for a substantial number of denoising steps and inherently restricts an effective global search of the solution space. To facilitate a more comprehensive exploration of the vast chemical space, it is imperative to consider not only the orientation toward the next state but also the final state. Herein, we define these two gradient vectors as the local gradient (t to $t-1$) and the global gradient (t to 0), respectively.

Karras *et al.* [33] demonstrated that the score-based diffusion generative framework can be implemented via a denoiser function:

$$\mathbb{E}_{y\sim p_{\text{data}}}\,\mathbb{E}_{n\sim\mathcal{N}(0,\sigma^2 I)}\,\|D(y+\epsilon;\sigma)-y\|_2^2,$$
$$\text{then}\quad \nabla_l \log p(l;\sigma) = \frac{D(l;\sigma)-l}{\sigma^2} = \frac{\hat{y}-l}{\sigma^2}, \tag{11}$$

where y and l are the ground true and noised samples, ϵ is noise, and D is the denoiser function (i.e. ϕ_θ in GlintDM).

From Equation 11, the global gradient, ∇l_g, can be written as

$$\nabla_{l_t} \log p(l_t;\sigma_t) \approx \hat{l}_0 - l_t =: \nabla l_g. \tag{12}$$

Using the closed-form posterior mean of DDPM:

$$\tilde{\mu}_t(l_t, l_0) = A_t\, l_0 + B_t\, l_t = l_{t-1} + \sqrt{\tilde{\beta}_t}\epsilon, \quad \epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I}) \tag{13}$$

$$A_t := \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}, \qquad B_t := \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}, \tag{14}$$

the local gradient, ∇l_l, can be written as

$$\begin{aligned}
\mathbb{E}[\nabla l_l \mid l_t, l_0] &= \mathbb{E}[l_{t-1} - l_t \mid l_t, l_0] && (15)\\
&= \tilde{\mu}_t(l_t, l_0) - l_t && (16)\\
&= A_t(l_0 - l_t) + (A_t + B_t - 1)\, l_t && (17)\\
&= A_t(l_0 - l_t) + r_t\, l_t && (18)\\
&\approx l_{t-1} - l_t =: \nabla l_l, && (19)
\end{aligned}$$

where $r_t := A_t + B_t - 1$. Replacing l_0 by its denoised estimate $\hat{l}_0$ gives

$$\begin{aligned}
\mathbb{E}[\nabla l_l \mid l_t, \hat{l}_0] &= A_t(\hat{l}_0 - l_t) + r_t\, l_t && (20)\\
&= A_t\,\nabla l_g + r_t\, l_t \xrightarrow[t\to 0]{} A_t\,\nabla l_g, && (21)
\end{aligned}$$

where, as $t\to 0$, $\beta_t\to 0$ and thus $r_t\to 0$, as illustrated in Fig. S1.

Therefore, combining global and local gradients yields a re-scaled ascent along the same score field:

$$\begin{aligned}
\lambda_t\nabla l_g + (1-\lambda_t)\nabla l_l &\approx \left[\lambda_t + (1-\lambda_t)A_t\right]\nabla l_g && (22)\\
&= \xi_t \nabla l_g && (23)\\
&\approx \xi_t \nabla_{l_t} \log p(l_t;\sigma_t), && (24)
\end{aligned}$$

where $\lambda_t = t/T$ and $\xi_t := \lambda_t + (1-\lambda_t)A_t$. This combination of the two gradients enables the diffusion model not only to follow the local stepwise transition but also to move globally beyond the immediate next step.

In our molecular diffusion model, the distributions of atom position and type, $[\mathbf{x}_t, \mathbf{h}_t]$, are regarded as the points in the two distributions. Given that $[\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0]$ and $[\mathbf{x}_{t-1}, \mathbf{h}_{t-1}]$ are estimated by the network ϕ_θ from $[\mathbf{x}_t, \mathbf{h}_t]$, we define global and local gradients by leveraging the finite difference method for $\mathbf{x}_t$ and reverse KL-divergence for $\mathbf{h}_t$:

$$\begin{aligned}
\nabla\mathbf{x}_l &= \mathbf{x}_{t-1} - \mathbf{x}_t, \quad \nabla\mathbf{x}_g = \hat{\mathbf{x}}_0 - \mathbf{x}_t\\
\nabla\mathbf{h}_l &= -\frac{\partial}{\partial\mathbf{h}_t} D_{KL}(\mathbf{h}_t\|\mathbf{h}_{t-1}) = -\exp(\mathbf{h}_t)(1+\mathbf{h}_t+\mathbf{h}_{t-1})\\
\nabla\mathbf{h}_g &= -\frac{\partial}{\partial\mathbf{h}_t} D_{KL}(\mathbf{h}_t\|\hat{\mathbf{h}}_0) = -\exp(\mathbf{h}_t)(1+\mathbf{h}_t+\hat{\mathbf{h}}_0), \qquad (25)
\end{aligned}$$

where $\nabla\mathbf{x}_l$ and $\nabla\mathbf{h}_l$ are local gradients, and $\nabla\mathbf{x}_g$ and $\nabla\mathbf{h}_g$ are global gradients. Specifically, the gradients of atom positions are derived from equations (12) and (19), which take a form similar to the finite difference method. For atom types, we compute gradients by differentiating the reverse KL-divergence, which is more suitable for categorical distributions. Lines 16 and 17 of Algorithm S2 correspond

Algorithm 3 Molecule generation procedure of GlintDM

Schematic overview. GlintDM generates ligands in three stages:
(1) *Position refinement*: repeating coordinate-only refinement to obtain a plausible pose
(2) *Candidate evaluation*: selecting the top-scoring candidate ligands
(3) *Ligand resampling*: regenerating final ligands with similar properties as candidates

Require: Binding site P, pretrained GlintDM
Initialize skip interval n, top-N size, buffers
Position refinement
Initialize $\mathbf{x}_T \sim \mathcal{N}(0, I)$
for each refinement step **do**
    Initialize $\mathbf{h}_T$ using Gumbel sampling
    Perform coordinate refinement via GlintDM.Sampling
    Update $\mathbf{x}_T \leftarrow \hat{\mathbf{x}}_0$
end for
Candidate evaluation
Extract top-N candidate ligands
Ligand resampling
for each candidate **do**
    Apply resampling noise to $\hat{\mathbf{h}}_0^i$
    Generate final $[\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0]$ using GlintDM.Sampling
    Append to NewLigands
end for
return NewLigands

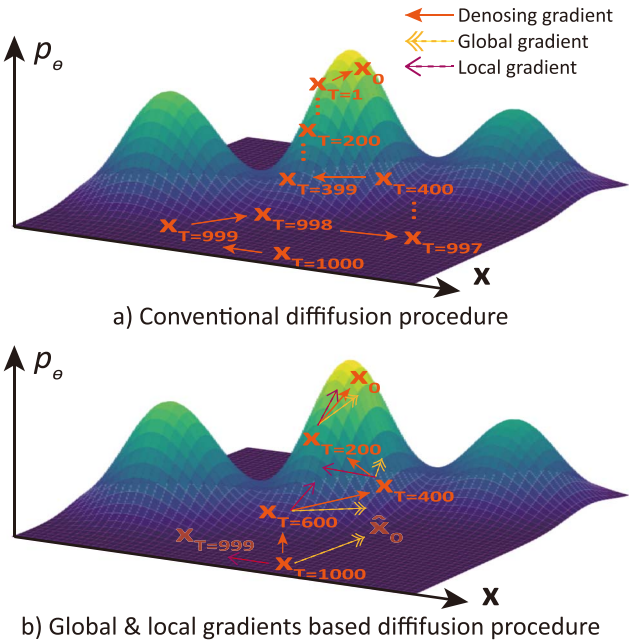

Figure 2 Global and local gradients-based denoising procedure, where only a few denoising transition steps are used to reach the optimal point.

to the combination of global and local gradients for atom positions and types, respectively. In the atom-type gradient, the global gradient is normalized to account for the large discrepancy between atoms at $T = t$ and $T = 0$.

Skip transition

In the standard denoising (reverse) process, a model performs step-by-step transitions from time step t to $t-1$, focusing solely on local noise removal between states X_t and X_{t-1}. However, when both global and local gradients are available—with the global gradients guiding the direction toward the final state and the local ones toward the next state—some intermediate transitions can be skipped. This enables the model to bypass redundant transitions, significantly reducing the number of denoising steps. GlintDM adopts this strategy by combining global and local gradients and adaptively adjusting the step size based on the current time step t (Fig. 2 and Algorithm 2 and S2). We refer to this global- and local-gradient-based transition mechanism as **skip transition**. The skip transition directly moves from time step t to $t-n$, and effectively skipping $n-1$ intermediate steps. The value of n thus controls the acceleration of the denoising process.

GlintDM

The skip transition mechanism significantly reduces the number of denoising steps, enabling efficient integration of multiple refinement and resampling procedures. To fully exploit these advantages, GlintDM is structured into three sequential phases: *position refinement*, *candidate evaluation*, and *ligand resampling* (Fig. 1 and Algorithm 3 and S3). In the *position refinement* phase, the model identifies the most probable binding site on the target protein. The *candidate evaluation* phase selects promising ligand candidates that satisfy multiple objectives. Finally, the *ligand resampling* phase generates new ligands that are both structurally stable and chemically similar to the selected candidates.

Position refinement

The goal of the position refinement phase is to identify the optimal binding sites. During this phase, multiple denoising processes are performed iteratively, where the output position generated at each iteration is reused as the input for the subsequent iteration. Specifically, in each iteration, the model uses only a few skip transitions for the fast denoising process (e.g. GlintDM performs 10 iterations with five skip transitions per iteration). As a result, while the initial iterations operate on highly noisy input positions, the inputs become progressively refined over time. This iterative refinement gradually shifts the model's goal from denoising in the early stages to accurately localizing binding sites in the later stages. Consequently, repeated diffusion procedures allow the model to find increasingly stable and precise binding sites.

Candidate evaluation

Although *position refinement* identifies probable binding sites, the resulting molecular properties ultimately depend on the ligand itself. In essence, discovering ligands satisfying target properties corresponds to identifying promising regions in the data distribution. Therefore, it is crucial to evaluate the ligands generated in the previous phase and select the most promising candidates. For that, we define a scoring function that accounts for multiple desirable molecular properties:

$$r(L|P) = \text{vina(L|P)} \times Min(QED(L), QED_{max}) \times \text{Min(SA(L),SA}_{\text{max}}) \quad (26)$$

where **vina**$(L|P)$ is the Vina score computed by AutoDock Vina [39], and **QED** and **SA** represent the quantitative estimate of drug-likeness [40] and synthetic accessibility [41], respectively. Inspired by the

geometric mean, the properties are combined multiplicatively rather than additively to yield a balanced overall score. The scoring function is used to select the most promising ligands as candidates.

Ligand resampling

In the ligand resampling phase, GlintDM generates diverse ligands while preserving the key properties of the candidates. When the atom-type probabilities and positions of the candidate ligands ($[\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_0]$) are used as input, noise is added only to $\hat{\mathbf{h}}_0$, while $\hat{\mathbf{x}}_0$ remains unperturbed. This enables the sampling of ligands with high molecular property values at stable positions. Note that the scoring function is not used during this phase to select ligands with higher scores.

Results

Datasets

We utilized two 3D protein–ligand datasets: CrossDocked [42] and Binding MOAD [43], following the preprocessing and data split protocols described by Guan *et al.* [9] and Schneuing *et al.* [10], respectively.

The original CrossDocked dataset contains ∼22.5 million docked protein–ligand pairs with varying binding quality. Guan *et al.* [9] removed complexes whose binding pose RMSD exceeded 1 Å, and clustered protein sequences at 30% sequence identity. From the resulting subset, 100 000 protein–ligand pairs were randomly selected for training, and 100 nonoverlapping proteins were reserved for testing (Table S2).

The Binding MOAD dataset [43] was curated following the procedure of Schneuing *et al.* [10], where ligands with QED > 0.3 were retained, atom types were restricted to **C**, **N**, **O**, **S**, **B**, **Br**, **Cl**, **P**, **I**, **F**, and complexes containing nonstandard amino acids were removed. To mitigate redundancy, no more than 50 ligands were randomly sampled per chemical component identifier (three-letter-code). After excluding corrupted entries, the resulting dataset comprised 40 344 training pairs, 246 validation pairs, and 130 testing pairs. In our experiment, we removed validation and test complexes that appeared in the CrossDocked and those with a Vina score <10. After this filtering, 144 pairs were retained as the final test set for the Binding MOAD benchmark (Table S3).

In both datasets, binding pockets were defined as the set of residues with at least one atom located within 10 Å of any ligand atom. The 100 000 training pairs from CrossDocked were used to train GlintDM, while the test sets of CrossDocked and Binding MOAD were used for its evaluation.

Evaluation

We evaluate GlintDM and baseline methods (details in "Benchmark models" of the supplementary material) for both binding affinity and following molecular properties: Vina Score, Vina Min, Vina Dock, and *High Affinity* (HA) for binding affinity evaluation and QED, SA, and diversity for molecular properties (details in "Metrics" of the supplementary material). Following Dorna *et al.* [25], we evaluate the trade-off between drug-likeness and binding affinity using the following thresholds: QED ≥ 0.4, SA ≥ 0.5, and Vina docking score ≤ –8.18 kcal/mol. The proportion of valid molecules meeting all three criteria is reported as the *Hit Rate* (HR).

Intermolecular interactions between the protein and the ligand may induce the ligand to adopt a specific folded conformation, which in turn can lead to structural instability, such as abnormally short or long interatomic distances. While most folded conformations in existing datasets are physically acceptable, 3D molecular generative models often produce unstable molecules that deviate from such conformations. To address this, most 3D generative models [9, 17, 23, 25, 26] employ `OpenBabel` [44] to construct complete molecules from generated atom types and coordinates. Although `OpenBabel` can refine unstable structures to some extent, it remains critical for generative models to directly generate structurally stable molecules without excessive reliance on post-processing.

Following the definitions proposed by Satorras *et al.* [45], *atom stability* is defined as the percentage of atoms maintaining chemically valid distances with neighboring atoms, whereas *molecule stability* refers to the percentage of molecules composed entirely of such stable atoms. Additionally, we define the *validity rate* as the percentage of generated molecules that can be successfully processed by Vina among all generated samples. This criterion is more stringent than a simple validity check, because some molecules, which successfully pass rdkit.Chem.SanitizeMol, may still fail to yield the Vina score.

A *steric clash* occurs when the distance between a protein atom and a ligand atom is shorter than the sum of their van der Waals radii, indicating an energetically unfavorable and physically implausible conformation. As reported by Harris *et al.* [46], diffusion-based generative models are prone to producing molecules with such steric clashes, although subsequent re-docking procedures can partially alleviate these issues.

Buttenschoen *et al.* [47] introduced *PoseBusters*, a validation framework designed to assess the physical plausibility of AI-generated protein–ligand docking poses. While primarily developed for docking evaluation, PoseBusters can also be applied to validate ligand structures generated for specific protein targets. The specific validation criteria are summarized in Table S1.

To assess geometric fidelity, we follow Guan *et al.* [9] and evaluate three key geometric properties against the reference dataset, CrossDocked: (i) the empirical distributions of all-atom and carbon–carbon bond distances, (ii) the Jensen–Shannon divergence (JSD) [48] between bond distance distributions, and (iii) the empirical distributions of ring sizes.

GlintDM

GlintDM requires just 70 denoising steps in total: 50 for position refinement—organized into 10 iterations with five skip transitions each—and 20 for ligand resampling. This is a remarkably small number compared with the baseline model, TargetDiff, which requires 1000 denoising steps.

We design two variants of GlintDM: *GlintDM-Mono* and *GlintDM-Multi*. In the candidate evaluation phase, *GlintDM-Mono* uses only the Vina Score as the scoring function, whereas *GlintDM-Multi* employs Equation 26 for multi-objective optimization. The former focuses on generating ligands with high binding affinity, while the latter aims to generate ligands that satisfy both high affinity and drug-likeness properties.

Hyperparameters of the diffusion network are the same as TargetDiff, and the additional hyperparameters are in Table S4. We trained GlintDM using a single NVIDIA GeForce RTX 3090 GPU without parallel processing. The experiments were conducted on a workstation equipped with dual Intel(R) Xeon(R) Gold 6230R CPUs (two sockets, 26 cores per socket, 104 threads in total) running at 2.10 GHz, and 502 GB of system memory.

Table 1 In CrossDocked, comparison of GlintDM with benchmark models across various metrics, where bold indicates the best performance, and HA and HR are high-affinity and hit rate, respectively. All benchmark results were reported as provided in the respective original papers, except for AR, Pocket2Mol, TargetDiff, and DecompDiff, whose results were obtained from [25]. "–" indicates no reported score in the original paper.

	Metrics	**Vina Score ↓**		**Vina Min ↓**		**Vina Dock ↓**		**QED ↑**	**SA ↑**	**Diversity ↑**	**HA (%) ↑**	**HR (%) ↑**
	Group name	**Mean**	**Med.**	**Mean**	**Med.**	**Mean**	**Med.**	**Mean**	**Mean**	**Mean**	**Mean**	**Rate**
Test set	CrossDocked	−6.36	−6.46	−6.71	−6.49	−7.45	−7.26	0.48	0.73	–	–	21
Non-Diff.	AR ([7])	−5.75	−5.64	−6.18	−5.88	−6.75	−6.62	0.51	0.65	0.70	37.9	12.9
	Pocket2Mol ([8])	−5.14	−4.70	−6.42	−5.82	−7.15	−6.77	0.56	**0.74**	0.69	48.4	24.3
Diff.	BindDM ([34])	−5.92	−6.87	−7.29	−7.34	−8.41	−8.37	0.51	0.58	0.68	64.8	–
	TargetDiff ([9])	−5.47	−6.30	−6.64	−7.04	−7.91	−7.96	0.48	0.58	0.72	58.1	20.5
	DecompDiff ([17])	−4.85	−6.03	−6.76	−7.09	−8.48	−8.50	0.44	0.59	0.63	64.4	24.9
	DECOMPOPT ([35])	−5.87	−6.81	−7.35	−7.72	−8.98	−9.01	0.48	0.65	0.60	73.5	–
	IPDiff ([26])	−6.42	−7.01	−7.45	−7.48	−8.57	−8.51	0.52	0.61	0.74	69.5	–
Diff. + guidance	TAGMOL ([25])	−7.02	−7.77	−7.95	−8.07	−8.59	−8.69	0.55	0.56	0.69	–	27.7
	UniGuide ([23])	−5.07	–	−6.62	–	−7.91	–	0.57	0.64	–	–	–
	BADGER ([24]) + TargetDiff	−7.70	−8.53	−8.33	−8.44	−8.91	−8.84	0.46	0.50	**0.78**	70.2	–
SDEs + RL	DiffAC ([36])	−9.07	−9.04	–	–	–	–	–	–	–	–	–
Flow matching	FlexSBDD ([37])	−6.64	−7.25	−8.27	−8.46	−9.12	−9.25	0.58	0.69	0.76	–	–
Bayesian + Flow Network	MolGRAFT-large ([38])	−6.61	−8.14	−8.14	−8.42	−9.25	−9.20	0.46	0.62	0.61	–	–
Ours	GlintDM-Mono	**−9.41**	**−9.52**	**−9.93**	**−9.98**	**−10.41**	**−10.47**	0.48	0.48	0.60	**92.0**	21.2
	GlintDM-Multi	−7.60	−7.84	−8.19	−8.24	−8.90	−9.03	**0.60**	0.60	0.68	75.4	**47.4**

Table 2 Comparison of GlintDM with baseline models on the Binding MOAD dataset across multiple metrics, where bold indicates the best performance. Metrics include validity rate, molecule stability (MS), atom stability (AS), high-affinity rate (HA), hit rate (HR), and time. All baseline results were reproduced using our implementation

	Metrics	**Validity ↑**	**MS ↑**	**AS ↑**	**Vina Score ↓**		**Vina Min ↓**		**QED ↑**	**SA ↑**	**Diversity ↑**	**HA (%) ↑**	**HR* (%) ↑**	**Time** (sec) ↓**
	Group name	**Rate**	**Rate**	**Rate**	**Mean**	**Med.**	**Mean**	**Med.**	**Mean**	**Mean**	**Mean**	**Mean**	**Rate**	**Mean**
Test set	Binding MOAD	–	0.446	0.943	−6.75	−6.67	−7.46	−7.43	0.53	**0.69**	–	–	25.0	–
Diff.	TargetDiff [9]	0.905	0.446	0.944	−6.01	−6.45	−6.93	−6.95	0.56	0.62	0.77	43.1	22.7	1240
	DiffSBDD [17]	0.849	0.143	0.737	12.73	0.68	−1.14	−3.72	0.46	0.61	**0.84**	3.2	1.0	**264**
Diff. + guidance	TAGMOL [25]	0.864	0.352	0.932	**−7.92**	−8.17	−8.52	−8.37	0.61	0.59	0.72	56.8	32.3	2463
Ours	GlintDM-Mono	0.990	0.490	0.960	−7.75	**−8.91**	**−9.09**	**−9.35**	0.59	0.55	0.66	**72.3**	34.2	**299**
	GlintDM-Multi	**0.993**	**0.540**	**0.961**	−7.19	−8.16	−8.37	−8.56	**0.65**	0.61	0.71	66.3	**47.5**	**299**

∗HR uses Vina min score instead of Vina docking score. ∗∗Time indicates the average time to generate molecules per one target pocket.

CrossDocked and binding MOAD

For the CrossDocked benchmark (Table 1), both variants of GlintDM achieved high binding affinity (Vina-based metrics) and multi-objective evaluation (hit rate) compared with other recently developed methods. Notably, *GlintDM-Mono* significantly outperforms existing benchmark methods across all Vina-based metrics. In particular, the hit rate of *GlintDM-Multi* reaches 47.4%, indicating that nearly half of the generated molecules satisfy strict thresholds for drug-likeness. This result highlights the effectiveness of GlintDM in multi-objective molecular generation.

In the Binding MOAD benchmark (Table 2), GlintDM also achieved high both binding affinity and hit rate compared with other methods. Remarkably, only GlintDM attains a validity rate exceeding 0.990, while all other benchmark models fail to surpass 0.91. In particular, although TAGMOL achieved a similar performance with GlintDM-Multi in Vina-related scores, it shows a relatively low validity rate (0.864), molecule stability (0.352), and atom stability (0.932). These results indicate that it failed to generate valid and stable molecules in some challenging test cases. Considering that the high hit rate reaches (47.5%), validity rate (0.993), molecule stability (0.540), and atom stability (0.961) of *GlintDM-Multi*, it demonstrates the ability to generate high-quality molecules even for challenging protein targets, outperforming competing models.

To assess the generality of GlintDM across diverse protein families, we summarized the evaluation metrics according to the PDB protein classification. Due to the large number of fine-grained PDB categories, we grouped them into broader major protein classes (Tables S2 and S3). As shown in Tables S5 and S6, GlintDM successfully generated appropriate molecules for all major classes except for the "Cell adhesion/invasion" category in the Binding MOAD dataset. Among proteins in the "Cell adhesion/invasion" category, PDB ID: 4F8L exhibits a highly unfavorable binding environment, where the reference ligand achieves a Vina score of only −1.077. Consequently, the generated molecules for the "Cell adhesion/invasion – 4F8L" target exhibit an appropriate Vina score (–3.913). These results demonstrate that GlintDM exhibits strong generalizability and robustness across a wide range of protein classes.

In addition, we compared the time required to generate molecules for a single target pocket across GlintDM and all benchmark methods. The batch size was set to 100 for all methods except TAGMOL, for which a batch size of 35 was used due to GPU memory constraints. As shown in Table 2, GlintDM generates molecules approximately four times faster than TargetDiff and eight times faster than TAGMOL. Specifically, on average, both GlintDM-Mono and Multi required 59 (± 16) seconds for the refinement phase, 215 (± 112) s for the candidate generation phase, and 25 (± 6) s for the candidate evaluation phase. These results indicate that although GlintDM itself is computationally efficient in generating molecules (84 s), the computation of Vina scores—highly dependent on the size and complexity of protein structures—remains the primary runtime bottleneck (215 s). In contrast, QED and SA calculations add negligible computational cost, resulting in nearly identical runtimes for GlintDM-Mono and GlintDM-Multi.

Table 3 Comparison of GlintDM and baseline models in terms of their ability to stably generate high-quality ligands in the CrossDocked. Metrics include molecule stability, atom stability, validity rate, average steric clashes, novelty, and scaffold novelty. The best scores are highlighted in bold

	AVG. molecule stability	AVG. atom stability	Validity rate	AVG. Steric Clashes	Novelty	Scaffold novelty
CrossDocked	0.430	0.937	–	8.28	–	–
GlintDM-Multi	**0.543**	**0.967**	**0.994**	**7.02**	0.985	0.951
GlintDM-Mono	0.457	0.966	0.965	9.56	**0.992**	**0.977**
TargetDiff	0.439	0.949	0.923	12.89	0.940	0.838
TAGMOL	0.354	0.936	0.92	9.73	0.968	0.915

Overall, GlintDM demonstrates an ability to rapidly and effectively explore regions of chemical space.

Ligand quality

To evaluate the ability of GlintDM to generate chemically and structurally stable ligands, we compared its performance with TargetDiff [9], TAGMOL [25], and the CrossDocked test set. As shown in Table 3, GlintDM variants generate molecules with higher stability than even the reference dataset, in terms of both atomic and molecular stability. To assess the model's distribution learning capability and to ensure that it does not merely memorize training examples, we evaluated both Novelty and Scaffold Novelty (details in the "Novelty" of the Supplementary file). Across both metrics, GlintDM achieves the highest degree of novelty among all compared methods. In addition, over 50% of the molecules generated by *GlintDM-Multi* are composed entirely of stable atoms and show lower steric clash scores on average than the reference dataset (Fig. S2).

We also used the PoseBusters framework [47] and three geometric metrics (atom-wise distance, bond-type distribution, and ring-size distribution) to evaluate structural plausibility against the Cross-Docked dataset. In the PoseBusters checklist (Table S7), *GlintDM-Multi* showed the most consistent scores with reference molecules. Moreover, across the three geometric properties, the GlintDM variants achieved the best or second-best performance (Tables S8 and S9 and Fig. S3). These results indicate that GlintDM (especially *Glint-Multi*) is capable of generating ligands that not only satisfy multi-objective properties but also exhibit high structural stability.

Ligand design for a human disease target

To assess the translational potential of GlintDM for the discovery of therapeutic ligands, we evaluated its performance on a clinically relevant target: the cystic fibrosis transmembrane conductance regulator (CFTR), whose genetic dysfunction underlies cystic fibrosis. Structural complexes of CFTR bound to GLPG1837 and Ivacaftor (PDB IDs: 6O1V and 6O2P, respectively), as reported by Liu, Fangyu, *et al.* [49], served as reference models. Using the protein structures from 6O1V and 6O2P, GlintDM generated two new compounds, referred to as *New Drug A* and *New Drug B* (details in "Ligand generation protocol for a human disease target" of the supplementary material). When analyzed using PyMOL [50], these compounds have more hydrogen bonds than reference drugs, as depicted in Fig. S4. Specifically, *New Drug A* established three hydrogen bonds with CFTR (residues TYR304, PHE305, and SER308), compared with two by GLPG1837 (TYR304 and SER308). Likewise, *New Drug B* engaged TYR304, GLY930, and PHE931 through three hydrogen bonds, whereas Ivacaftor formed a single hydrogen bond. In addition, the GlintDM-generated ligands exhibited favorable QED and SA scores, and lower Vina binding scores than the reference drugs (Fig. S4). Taken together, these findings indicate that the GlintDM-generated compounds may serve as promising candidates for therapeutic development.

Ligand design in noncanonical conditions

Most existing drug generative models have primarily focused on predefined canonical ligand-binding pockets, which are orthosteric binding sites derived from holo protein structures. However, practical drug discovery often involves noncanonical conditions, such as targets with only an apo structure or with an allosteric pocket, which is a spatially distinct regulatory site separate from the main binding pocket. To evaluate the robustness of GlintDM under these non-canonical conditions, we applied our model to two challenging cases: allosteric pockets and apo protein structures.

Allosteric pockets

The ASBench Core-diversity set [51] consists of 147 structurally diverse allosteric sites. Among these, only 75 complexes satisfied the input requirements of AutoDock Vina and exhibited valid binding scores (Vina scores lower than 1000). Consequently, we used 75 allosteric pockets for evaluation (Table S11), and GlintDM generated 100 ligands for each allosteric pocket.

In Table 4, the Vina scores of ASBench (mean: –1.21) were significantly higher than those of the CrossDocked (–6.71) and Binding MOAD (–6.75) datasets, indicating that ASBench constitutes a substantially harder benchmark. Due to this increased difficulty, both GlintDM variants and TargetDiff produced molecules with relatively lower molecular stability in ASBench compared with CrossDocked and Binding MOAD. Nevertheless, the GlintDM variants consistently achieved lower Vina-related scores while maintaining reasonable QED and SA values. Moreover, GlintDM variants outperformed TargetDiff in terms of validity rate, high-affinity, and hit rate. Together, these results demonstrate the robustness of GlintDM even in allosteric pockets.

Apo structures

To evaluate the performance of GlintDM on apo protein structures, we utilized the BindingDB dataset [52], which provides experimentally measured protein-ligand pairs with reported pIC_{50} values, even in the absence of holo structures. Following the categorization established by Karimi *et al.* [53], proteins were grouped into four classes: ion channel, G-protein-coupled receptor (GPCR), nuclear estrogen receptor (ER), and kinase.

Since these apo protein–ligand pairs lack binding site annotations, we first needed to predict putative pockets on the apo structures. To this end, we constructed two experimental settings using DeepSurf

Table 4 Comparison of GlintDM with TargetDiff on the ASBench dataset across multiple metrics including validity rate, molecule stability (MS), atom stability (AS), high-affinity rate (HA), QED, SA, diversity, high-affinity rate (HA), and hit rate (HR)

	Metrics	Validity ↑	MS ↑	AS ↑	Vina Score ↓		Vina Min ↓		QED ↑	SA ↑	Diversity ↑	HA (%) ↑	HR* (%) ↑
	Cases	Rate	Mean	Mean	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Mean	Rate
Test set	ASBench [51]	–	0.416	0.937	−1.208	0.0	−2.267	0.0	0.45	0.68	–	–	6.6
Diff.	TargetDiff	0.809	0.299	0.930	−4.93	−6.18	−6.18	−6.88	0.49	0.59	0.71	62.7	17.0
Ours	GlintDM-Mono	0.984	0.246	0.923	−7.89	−8.38	−9.07	−9.00	0.50	0.53	0.64	93.1	24.5
	GlintDM-Multi	0.991	0.280	0.921	−7.03	−7.63	−8.09	−8.20	0.59	0.60	0.69	89.3	38.4

*HR uses Vina min score instead of Vina docking score.

Table 5 Performance comparison of GlintDM-Multi across four apo protein classes using two binding-site prediction methods, DeepSurf and AF3. The Channel, GPCR, and Kinase classes each contain ten reference ligand–protein complexes, whereas the ER class includes only a single complex

	Metrics	Vina Score ↓		Vina Min ↓		QED ↑		SA ↑	
Prot Class	Binding site	Ref. ligands [Min, Avg, Max]	GlintDM-Multi [Min, Avg, Max]	Ref. ligands [Min, Avg, Max]	GlintDM-Multi [Min, Avg, Max]	Ref. ligands [Min, Avg, Max]	GlintDM-Multi [Min, Avg, Max]	Ref. ligands [Min, Avg, Max]	GlintDM-Multi [Min, Avg, Max]
Channel	DeepSurf	[−9.1, 1.5, 42.5]	[−13.9, −4.24, 70.7]	[−9.1, −7.1, −2.6]	[−14.8, −6.5, 70.7]	[0.45, 0.7, 0.91]	[0.02, 0.57, 0.93]	[0.58, 0.75, 0.86]	[0.29, 0.57, 0.97]
	AF3	[−11.5, −5.8, 9.7]	[−15.7, −8.4, −0.8]	[−12.6, −7.0, 6.5]	[−16.0, −8.7, −1.0]		[0.1, 0.64, 0.92]		[0.31, 0.61, 0.9]
ER	DeepSurf	−9.4	[−11.5, −9.4, −7.1]	−9.4	[−11.6, −9.9, −7.5]	0.52	[0.23, 0.71, 0.92]	0.91	[0.45, 0.61, 0.9]
	AF3	−8.6	[−12.7, −10.6, −8.3]	−9.9	[−13.0, −11.0, −9.0]		[0.49, 0.72, 0.90]		[0.46, 0.59, 0.91]
GPCR	DeepSurf	[−12.7, −6.9, 15.2]	[−14.7, −8.1, −2.8]	[−12.7, −7.2, 15.1]	[−15.3, −8.5, −3.3]	[0.23, 0.45, 0.72]	[0.06, 0.53, 0.94]	[0.51, 0.69, 0.82]	[0.22, 0.57, 0.85]
	AF3	[−13.5, −6.9, 19.9]	[−16.1, −9.02, −4.1]	[−14.0, −10.2, −4.9]	[−16.2, −9.4, −4.6]		[0.09, 0.56, 0.92]		[0.22, 0.56, 0.92]
Kinase	DeepSurf	[−9.5, −6.7, 3.0]	[−14.0, −7.6, −3.4]	[−9.7, −7.6, −5.4]	[−14.4, −8.0, −3.6]	[0.24, 0.61, 0.88]	[0.16, 0.63, 0.93]	[0.51, 0.64, 0.76]	[0.29, 0.57, 0.86]
	AF3	[−10.4, −5.5, −1.8]	[−14.9, −1.7, 77.3]	[−12.1, −7.2, −3.9]	[−15.5, −4.0, 73.5]		[0.1, 0.59, 0.94]		[0.27, 0.6, 0.91]

[54] and AlphaFold3 (AF3) [55]: (i) apo structures with binding sites predicted by DeepSurf, and (ii) AF3-predicted holo structures. Specifically, DeepSurf identifies candidate pockets directly on apo proteins, whereas AF3 generates putative holo conformations of protein-ligand complexes. Accordingly, we applied GlintDM to both the apo structures with predicted binding sites and the AF3-predicted holo structures.

From the available apo structures, we selected ten distinct proteins with the highest pIC_{50} values each for the ion channel, GPCR, and kinase classes, and included a single ER protein due to the limited availability of apo structures in that category (Table S11). For each selected protein, we retrieved the top 25 protein–ligand pairs with the highest pIC_{50} values from BindingDB and computed their Vina binding scores for both the DeepSurf-predicted binding sites and the AF3-predicted holo conformations. Among the 25 candidate ligands, we then selected one ligand that exhibited low Vina scores in both settings as the reference ligand for that protein. For evaluation, we used the two pockets generated by DeepSurf and AF3 for each reference protein–ligand pair, yielding a total of 62 pockets (31 proteins × 2 pocket types), and GlintDM generated 100 ligand candidates for each pocket.

As summarized in Table 5, some protein–ligand complexes exhibited high Vina scores in both the reference structures and the molecules generated by GlintDM, indicating that the binding prediction models had selected incorrect pocket sites for these cases. Apart from these exceptions, GlintDM successfully generated ligands with lower Vina scores while maintaining favorable QED and SA scores across all protein classes.

Ablation study

Skip transition

To evaluate the capability of the skip transition to generate stable molecules with a limited number of denoising steps, we compared the outputs of GlintDM using 10, 20, and 50 skip transitions, without applying the three phases. As shown in Table 6, GlintDM begins to produce chemically valid molecules when the number of skip transitions exceeds 20.

To assess whether a molecular diffusion model can benefit from the core components of GlintDM in the absence of the skip transition, we incorporated these components into TargetDiff under the following configurations: (i) TargetDiff using a few standard reverse (denoising) steps; (ii) TargetDiff with refinement; (iii) TargetDiff with candidate evaluation and resampling. In detail, TargetDiff *w/* 10 pos-refine used a total of 50 reverse steps across 10 iterations for position refinement. By contrast, TargetDiff *w/* 1000 Rev. steps + Cand. Eval. + Resampl. employed 1000 reverse steps to generate candidates, followed by an additional 1000 reverse steps for resampling, resulting in 2000 reverse steps in total. As shown in Table 6, TargetDiff mostly fails to generate valid molecules and is unable to effectively utilize the core components without the skip transition.

Additionally, we investigated whether *TargetDiff* could leverage the skip transition-based resampling (i.e. the resampling scheme of *GlintDM*). To this end, molecules generated by *TargetDiff* using 1000 reverse steps were applied to the candidate evaluation and resampling phases of *GlintDM*. This variant differs from *GlintDM-Multi* only in using 1000 reverse steps instead of 10 position refinements. Unlike *TargetDiff* with candidate evaluation and resampling, this version exhibited improved performance across all metrics except for diversity. However, its metrics, except for the validity rate, were lower than those of *GlintDM-Multi*. These results highlight the efficiency of the position refinement and the skip transition of GlintDM.

Position refinement

We hypothesize that repeating the denoising process increases the likelihood of discovering structurally stable and potentially favorable binding sites. To validate this hypothesis, we compared the results of the ligand generation without the position refinement phase (50 skip transitions) and with five and 10 position refinements, while keeping the total number of denoising steps fixed at 50. During the resampling phase, we used 20 skip transitions in GlintDM variants. As shown

Table 6 Ablation study on the components and configurations of GlintDM using Equation 26 in the CrossDocked dataset. "TargetDiff variants" refer to configurations where TargetDiff is applied to key modules of GlintDM. Metrics include validity rate, molecule stability (MS), atom stability (AS), Vina score, QED, SA, and diversity. Abbreviations: Skip-Trans. = skip transition, Rev. = reverse, Cand. = candidate, Eval. = evaluation, Resampl. = resampling

	Metrics	**Validity ↑**	**MS ↑**	**AS ↑**	**Vina Score ↓**		**Vina Min ↓**		**QED ↑**	**SA ↑**	**Diversity ↑**
	Cases	**Rate**	**Mean**	**Mean**	**Mean**	**Med.**	**Mean**	**Med.**	**Mean**	**Med.**	**Mean**
Using only Skip transition	10 Skip-Trans.	0.7237	0.001	0.306	−4.71	−5.42	−5.72	−5.73	0.38	0.60	0.76
	20 Skip-Trans.	0.9160	0.012	0.558	−6.02	−6.53	−6.81	−6.76	0.44	0.57	0.76
	50 Skip-Trans.	0.9596	0.246	0.898	−5.63	−6.41	−6.41	−6.81	0.50	0.57	0.74
TargetDiff variants	using 50 Rev. steps	0.1152	0.001	0.343	0.17	−2.40	−2.91	−3.32	0.25	0.39	0.84
	using 100 Rev. steps	0.1230	0.001	0.425	−1.18	−2.54	−3.09	−3.41	0.26	0.41	0.86
	w/ 10 pos-refine (50 Rev. steps)	0.5569	0.087	0.859	−4.16	−4.86	−5.22	−5.23	0.32	0.49	0.78
	w/ 1000 Rev. steps + Cand. Eval. + Resamp. (1000 Rev. steps)	0.8428	0.431	0.953	−6.01	−6.20	−6.71	−6.69	0.50	0.63	0.76
	w/ 1000 Rev. steps + Cand. Eval. + Resamp. of GlintDM (20 Skip-Trans.)	0.9968	0.4751	0.960	−7.54	−7.57	−8.03	−7.97	0.56	0.59	0.67
GlintDM variants	10 pos-refine + Cand. Eval. + Resamp. (**GlintDM-Multi**)	0.9949	0.544	0.968	−7.60	−7.84	−8.19	−8.24	0.60	0.60	0.68
	5 pos-refine + Cand. Eval. + Resamp.	0.9967	0.489	0.960	−7.68	−7.90	−8.31	−8.35	0.61	0.59	0.66
	1 pos-refine + Cand. Eval. + Resamp.	0.9958	0.395	0.943	−7.52	−7.64	−8.15	−8.13	0.59	0.58	0.67
	10 pos-refine + Resamp.	0.9483	0.505	0.962	−5.70	−6.68	−7.00	−7.17	0.52	0.59	0.73
	GlintDM-Multi *w/o* local gradient	0.9970	0.510	0.962	−7.45	−7.60	−8.05	−8.08	0.60	0.60	0.68
	GlintDM-Multi *w/o* global gradient	0.1268	0.003	0.183	22.45	11.96	3.52	−2.72	0.30	0.37	0.78

in Table 6, "10 pos-refine + Cond. Eval. + Resamp". (GlintDM-Multi) produced the most stable ligands and achieved high overall molecular properties among the three cases. Despite all settings using the same number of total denoising steps, distributing the process into multiple shorter refinement phases resulted in better performance.

Fig. S5 visually illustrates the progression of atom stability and Vina scores across iterations. Ligands shown in light red were generated during the early iterations, with the color gradually darkening as iterations progressed. The yellow ligand represents the final output generated at the last iteration. As the number of iterations increases, the atoms in the red ligands progressively move toward their correct positions, eventually resembling the final yellow ligand. Furthermore, as shown in Fig. S6, increasing the number of iterations results in structurally stable ligands with lower Vina scores. In addition, the centroid positions of the refined molecules are closely aligned with those of the native ligands, as reported in Table S12.

As illustrated in Fig. S7, the t-SNE plots of the molecular embeddings generated by GlintDM clarify how the refinement process progressively improves the generated molecules across iterations. Molecules produced during the early stages cluster within a compact region of the embedding space, indicating that the dual ligand–pocket embeddings initially capture only coarse, non-specific structural patterns with low chemical validity and weak site complementarity. As iterations progress, the embeddings gradually migrate toward distinct regions associated with higher atom stability and improved affinity scores. Notably, this directional movement reflects the model's increasing ability to yield chemically meaningful and pocket-consistent structures as the refinement steps proceed.

Candidate evaluation

While the position refinement phase aims to identify the optimal binding site, the candidate evaluation phase is designed to select ligand candidates satisfying the desired target properties. As anticipated, "10 pos-refine + Resamp" tends to generate molecules with lower target property scores compared with those generated with all three phases (Table 6).

Global and local gradients

To evaluate the contribution of global and local gradients, we compared the full gradient setting with a variant using only global and local gradients. "GlintDM-Multi *w/o* local gradient" exhibited lower molecular stability and reduced Vina-related scores. This indicates the utility of local gradients in the skip transition, whereas $T \rightarrow 0$, the influence of global gradients decreases, and the local gradient becomes increasingly dominant, guiding a more cautious transition and generating more stable and reliable samples. In contrast, "GlintDM-Multi *w/o* global gradient" was unable to generate valid molecules in a reliable manner, indicating that the skip transition cannot operate without the global gradient.

Hyperparameter setting

GlintDM would make different results according to the choices of hyperparameters. First, increasing the number of refinement iterations or skip transitions generally improves structural stability; however, we found that 10 refinement iterations with a skip interval of 5 provide a good balance between accuracy and efficiency. Second, using a smaller number of top-ranked candidate ligands during the candidate evaluation stage can increase target objective scores but may reduce molecular diversity. In contrast, generating a larger pool of initial samples during position refinement enhances both objective scores and diversity, although this comes with a higher computational cost, particularly due to Vina score calculations. Notably, the QED and SA components of the scoring function can be adjusted flexibly, as their computation is negligible. Finally, in the ligand resampling stage, the skip interval has minimal influence on the final outputs. Nonetheless, using more than or equal to 10 skip transitions is recommended to ensure stable generation of new ligands.

Discussion

Through evaluations on two benchmark datasets and comprehensive ablation studies, we demonstrate that GlintDM can rapidly generate ligands that satisfy multiple target-specific objectives. In particular, the introduction of skip transitions enables the use of multiple short

diffusion processes. Despite these strengths, GlintDM has certain limitations. First, it relies on external evaluation tools, such as RDKit [56] and AutoDock Vina [39], to guide the sampling process toward promising regions of the data distribution. Second, the model exhibits limited diversity in generated molecules. This is primarily due to the task's focus on a specific target protein, as well as the resampling phase, which tends to generate molecules that are structurally and chemically similar to the selected candidates. Third, similar to other benchmark methods, our model also assumes protein pockets to be stationary, which does not fully reflect the dynamic nature of protein–ligand interactions.

Although DDIM [18] and CMs [19] also accelerate sampling through deterministic updates, they do not leverage probabilistic benefits such as stochastic robustness and expressive exploration. In contrast, GlintDM integrates the strengths of both deterministic and probabilistic diffusion frameworks: global gradients are computed deterministically from the predicted x_0, while local gradients follow a probabilistic reverse process. This hybrid design substantially reduces the number of transition steps while producing stable multi-objective molecules as well as higher novelty and scaffold novelty.

Key Points

- We develop the **skip transition**, a novel denoising procedure based on global and local gradients, which significantly reduces intermediate denoising steps.
- We develop **GlintDM**, a skip transition-based drug generative framework consisting of three phases: position refinement, candidate evaluation, and ligand resampling.
- Our method is capable of generating stable and multi-objective molecules for target proteins using only 70 denoising steps.
- Our method outperformed other recently developed methods on the CrossDocked and Binding MOAD datasets for Vina-related scores.

Author contributions

Sejin Park: Formal analysis, Software, Data curation, Validation, Writing—original draft, Writing—review & editing; Minjae Chung: Data curation, Writing—review & editing; Hyunju Lee: Conceptualization, Formal analysis, Methodology, Validation, Writing—original draft, Writing—review & editing.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflicts of interest

There are no competing interests.

Funding

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2025-00519063), an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2019-0-00567, Development of Intelligent SW Systems for Uncovering Genetic Variation and Developing Personalized Medicine for Cancer Patients with Unknown Molecular Genetic Mechanisms and No.2019-0-01842, Artificial Intelligence Graduate School Program (GIST)), and InnoCORE program of the Ministry of Science and ICT(GIST InnoCORE KH0860).

Data availability

The source code and dataset for GlintDM are available on GitHub: https://github.com/DMCB-GIST/GlintDM.

References

1. Lyu J, Wang S, Balius TE. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* 2019;**566**:224–9. https://doi.org/10.1038/s41586-019-0917-9
2. Tang Y, Moretti R, Meiler J. Recent advances in automated structure-based de novo drug design. *J Chem Inf Model* 2024;**64**:1794–805. https://doi.org/10.1021/acs.jcim.4c00247
3. Ahn S, Kim J, Lee H, *et al.* Guiding deep molecular optimization with genetic exploration. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, (eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020).* Red Hook, NY: Curran Associates, Inc., 2020, 12008–21.
4. Park S, Lee H. A molecular generative model with genetic algorithm and tree search for cancer samples. arXiv 2021; arXiv:2112.08959.
5. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 1988;**28**:31–6. https://doi.org/10.1021/ci00057a005
6. Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization. arXiv 2014; arXiv:1409.2329.
7. Luo S, Guan J, Ma J, *et al.* A 3D generative model for structure-based drug design. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang P, Vaughan JW, (eds.), *Advances in Neural Information Processing Systems 34 (NeurIPS 2021).* Red Hook, NY: Curran Associates, Inc., 2021, 6229–39.
8. Peng X, Luo S, Guan J, *et al.* Pocket2Mol: efficient molecular sampling based on 3D protein pockets. In: Chaudhuri K, Jegelka S, (eds.), *Proceedings of the 39th International Conference on Machine Learning (ICML 2022).* PMLR, 2022, 17644–55.
9. Guan J, Qian WW, Peng X, *et al.* 3D equivariant diffusion for target-aware molecule generation and affinity prediction. In: *Proceedings of the 11th International Conference on Learning Representations (ICLR 2023).* 2023.
10. Schneuing A, Harris C, Yuanqi D. *et al.* Structure-based drug design with equivariant diffusion models. *Nat Comput Sci* 2024;**4**:899–909. https://doi.org/10.1038/s43588-024-00737-x
11. Yoo K, Oertell O, Lee J *et al.* TurboHopp: Accelerated molecule scaffold hopping with consistency models. In: Ranzato MA, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 37 (NeurIPS 2024).* Red Hook, NY, USA: Neural Information Processing Systems Foundation, Inc., 2024, 41157–85.
12. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin H, (eds.), Advances in Neural Information Processing Systems 33 (NeurIPS 2020).Curran Associates, Inc., Red Hook, NY, 2020:6840–51.
13. Yang S, Sohl-Dickstein J, Kingma DP, *et al.* Score-based generative modeling through stochastic differential equations. In: *Proceedings*

of the 9th International Conference on Learning Representations (ICLR 2021). 2021.
14. Huang L, Zhang H, Xu T, *et al.* MDM: molecular diffusion model for 3D molecule generation. In: *Proceedings of the AAAI Conference on Artificial Intelligence.* Palo Alto, CA: AAAI Press, 2023;37:5105–12. https://doi.org/10.1609/aaai.v37i4.25639
15. Huang L, Tingyang X, Yang Y. *et al.* A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat Commun* 2024;**15**:2657. https://doi.org/10.1038/s41467-024-46569-1
16. Torge J, Harris C, Mathis SV, *et al.* DiffHopp: a graph diffusion model for novel drug design via scaffold hopping. In: *Proceedings of the 2023 ICML Workshop on Computational Biology.* HI, USA: Honolulu, 2023.
17. *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, Hawaii, USA: PMLR 202, 2023. Copyright 2023 by the author(s).
18. Song J, Meng C, Ermon S. Denoising diffusion implicit models. In: *Proceedings of the 9th International Conference on Learning Representations (ICLR 2021).* 2021.
19. Yang S, Dhariwal P, Chen M, *et al.* Consistency models. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 36 (NeurIPS 2023).* Red Hook, NY: Curran Associates, Inc., 2023.
20. Satorras VG, Hoogeboom E, Welling M. E(n) equivariant graph neural networks. In: Meila M, Zhang T, (eds.), *Proceedings of the 38th International Conference on Machine Learning (ICML 2021).* Proceedings of Machine Learning Research (PMLR), vol. 139, 2021, 9323–9332.
21. Fuchs F, Worrall D, Fischer V, *et al.* SE(3)-transformers: 3D roto-translation equivariant attention networks. In: Larochelle H, Ranzato M, Hadsell R, *et al.* (eds.), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020).* Red Hook, NY: Curran Associates, Inc., 2020, 1970–81.
22. Hoogeboom E, Satorras VG, Vignac C, *et al.* Equivariant diffusion for molecule generation in 3D. In: Chaudhuri K, Jegelka S, (eds.), *Proceedings of the 39th International Conference on Machine Learning (ICML 2022).* Baltimore, MD, USA: Proceedings of Machine Learning Research (PMLR), vol. 162, 2022, 8867–87.
23. Ayadi S, Hetzel L, Sommer J, *et al.* Unified guidance for geometry-conditioned molecular generation. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 37 (NeurIPS 2024).* Red Hook, NY: Curran Associates, Inc., 2024, 138891–924.
24. Jian Y, Wu C, Reidenbach D, Krishnapriyan AS. General binding affinity guidance for diffusion models in structure-based drug design. *J Chem Inf Model* 2026. https://doi.org/10.1021/acs.jcim.5c01166
25. Dorna V, Subhalingam D, Kolluru K, *et al.* TAGMol: target-aware gradient-guided molecule generation. In: *Proceedings of the 1st Machine Learning for Life and Material Sciences Workshop at ICML 2024.* 2024.
26. Huang Z, Yang L, Zhou X, *et al.* Protein–ligand interaction prior for binding-aware 3D molecule diffusion models. In: *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024).* 2024.
27. Lugmayr A, Danelljan M, Romero A, *et al.* RePaint: inpainting using denoising diffusion probabilistic models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2022).* Los Alamitos, CA, USA: IEEE Computer Society, 2022, 11461–71.
28. Trippe BL, Yim J, Tischer D, *et al.* Diffusion probabilistic modeling of protein backbones in 3D for the motif-scaffolding problem. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 35 (NeurIPS 2022).* Red Hook, NY: Curran Associates, Inc., 2022.
29. Zhang Z, Shen WX, Liu Q, *et al.* Efficient generation of protein pockets with PocketGen. *Nat Mach Intell* 2024;**6**:1382–95. https://doi.org/10.1038/s42256-024-00920-9
30. Zhang Z, Li Z, Huang Z, *et al.* Full-atom protein pocket design via iterative refinement. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 36 (NeurIPS 2023).* Red Hook, NY: Curran Associates, Inc., 2023, 16816–36.
31. Cao D, Chen M, Zhang R, *et al.* SurfDock is a surface-informed diffusion generative model for reliable and accurate protein–ligand complex prediction. *Nat Methods* 2025;**22**:310–22. https://doi.org/10.1038/s41592-024-02516-y.
32. Jang E, Gu S, Poole B. Categorical reparameterization with Gumbel-Softmax. In: *Proceedings of the 5th International Conference on Learning Representations (ICLR 2017).* 2017.
33. Karras T, Aittala M, Aila T, *et al.* Elucidating the design space of diffusion-based generative models. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 35 (NeurIPS 2022).* Red Hook, NY: Curran Associates, Inc., 2022, 26565–77.
34. Huang Z, Yang L, Zhang Z, *et al.* Binding-adaptive diffusion models for structure-based drug design. In: *Proceedings of the AAAI Conference on Artificial Intelligence.* Palo Alto, CA: AAAI Press, 2024, 38, 12671–9. https://doi.org/10.1609/aaai.v38i11.29162
35. Zhou X, Cheng X, Yang Y, *et al.* DecompOpt: controllable and decomposed diffusion models for structure-based molecular optimization. In: *Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024).* 2024.
36. Zhou X, Wang L, Zhou Y. Stabilizing policy gradients for stochastic differential equations via consistency with perturbation process. arXiv 2024; arXiv:2403.04154.
37. Zhang Z, Wang M, Liu Q. FlexSBDD: structure-based drug design with flexible protein modeling. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 38 (NeurIPS 2024).* Red Hook, NY: Curran Associates, Inc., 2024.
38. Yanru Q, Qiu K, Song Y, *et al.* MolCRAFT: structure-based drug design in continuous parameter space. In: *Proceedings of the 41st International Conference on Machine Learning (ICML 2024).* Vienna, Austria: Proceedings of Machine Learning Research (PMLR), vol. 235, 2024.
39. Eberhardt J, Santos-Martins D, Tillack AF. *et al.* AutoDock Vina 1.2. 0: new docking methods, expanded force field, and python bindings. *J Chem Inf Model* 2021;**61**:3891–8. https://doi.org/10.1021/acs.jcim.1c00203
40. Richard Bickerton G, Paolini GV, Besnard J. *et al.* Quantifying the chemical beauty of drugs. *Nat Chem* 2012;**4**:90–8. https://doi.org/10.1038/nchem.1243
41. Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Chem* 2009;**1**:1–11. https://doi.org/10.1186/1758-2946-1-8
42. Francoeur PG, Masuda T, Sunseri J. *et al.* Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J Chem Inf Model* 2020;**60**:4200–15. https://doi.org/10.1021/acs.jcim.0c00411

43. Liegi H, Benson ML, Smith RD. *et al.* Binding MOAD (mother of all databases). *Proteins* 2005;**60**:333–40. https://doi.org/10.1002/prot.20512
44. O'Boyle NM, Banck M, James CA. *et al.* Open Babel: an open chemical toolbox. *J Chem* 2011;**3**:1–14.
45. Satorras VG, Hoogeboom E, Fuchs F, *et al.* E(n) equivariant normalizing flows. In: Ranzato M, Beygelzimer A, Dauphin Y, *et al.* (eds.), *Advances in Neural Information Processing Systems 34 (NeurIPS 2021).* Red Hook, NY: Curran Associates, Inc., 2021, 4181–92.
46. Harris C, Didi K, Jamasb AR, *et al.* PoseCheck: generative models for 3D structure-based drug design produce unrealistic poses. In: *Proceedings of the Machine Learning for Structural Biology (MLSB) Workshop at NeurIPS 2023.* (GenBio@NeurIPS 2023), 2023.
47. Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem Sci* 2024;**15**:3130–9. https://doi.org/10.1039/d3sc04185a
48. Lin J. Divergence measures based on the Shannon entropy. *IEEE Trans Inf Theory* 1991;**37**:145–51. https://doi.org/10.1109/18.61115
49. Liu F, Zhang Z, Levit A. *et al.* Structural identification of a hotspot on CFTR for potentiation. *Science* 2019;**364**:1184–8. https://doi.org/10.1126/science.aaw7611
50. DeLano WL. *et al.* PyMOL: an open-source molecular graphics tool. *CCP4 Newsl Protein Crystallogr* 2002;**40**:82–92.
51. Huang W, Wang G, Shen Q. *et al.* ASBench: benchmarking sets for allosteric discovery. *Bioinformatics* 2015;**31**:2598–600. https://doi.org/10.1093/bioinformatics/btv169
52. Liu T, Lin Y, Wen X. *et al.* BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res* 2007;**35**:D198–201.
53. Karimi M, Di W, Wang Z. *et al.* DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* 2019;**35**:3329–38. https://doi.org/10.1093/bioinformatics/btz111
54. Mylonas SK, Axenopoulos A, Daras P. DeepSurf: a surface-based deep learning approach for the prediction of ligand binding sites on proteins. *Bioinformatics* 2021;**37**:1681–90. https://doi.org/10.1093/bioinformatics/btab009
55. Abramson J, Adler J, Dunger J. *et al.* Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* 2024;**630**:493–500. https://doi.org/10.1038/s41586-024-07487-w
56. Landrum G. *Rdkit: Open-Source Cheminformatics Software* 2016.

Briefings in Bioinformatics, 2026, **27**, bbag033
https://doi.org/10.1093/bib/bbag033
Problem Solving Protocol