TheSelective: Dual Affinity–Guided Diffusion for Selective Molecular Generation

Hyoungjoon Park[1], Hwanhee Kim[1], Seungyeon Choi[1], Seungyong Lee[1], Yoonju Kim[1], and Sanghyun Park[1†]

[1] Yonsei University, Seoul, Republic of Korea
{ktori1361,hwanhee,tmddus1553,seungyong,yooonju49,sanghyun}@yonsei.ac.kr

Abstract. Diffusion-based generative models have shown promise in Structure-Based Drug Design (SBDD) by generating 3D ligands with strong binding affinities. However, existing models primarily focus on maximizing absolute potency toward a single target, neglecting selectivity: the ability to maximize the affinity gap between a target and potential off-targets. To address this, we propose **TheSelective**, a novel framework that enhances molecular selectivity via an asymmetric dual-guidance mechanism. Our approach captures on-target interactions through complex-based graphs while estimating off-target affinity via cross-attention between independent protein and ligand embeddings, bypassing the need for off-target docked structures. By jointly applying these dual-affinity signals to both atom types and spatial coordinates, TheSelective steers the generative process toward a maximized affinity gap across both chemical and geometric spaces. Experiments on the CrossDocked2020 dataset demonstrate that our model generates ligands with superior selectivity against both structurally similar and dissimilar off-targets. We validate the effectiveness of our framework through comprehensive ablation studies and case analyses, highlighting its potential for real-world drug discovery. The code is available at https://github.com/DannyJPark/TheSelective

Keywords: Selectivity · Binding Affinity · Guidance · Diffusion Model.

1 Introduction

Structure-based drug design (SBDD) has emerged as a cornerstone of modern pharmaceutical development, leveraging three-dimensional structural information of target proteins to rationally design therapeutic molecules with precise geometric fit and chemical complementarity. Recent advances in deep generative models, particularly diffusion-based approaches, have demonstrated remarkable capabilities in learning the complex distributions of protein-ligand interactions. Models such as TargetDiff [7], DecompDiff [8], and Pocket2Mol [15] have shown success in generating 3D molecules that satisfy the structural constraints of specific protein pockets. Building upon these foundations, guidance-based methods like KGDiff [16], BADGER [11], and TAGMol [4] have further improved binding

† Corresponding author: sanghyun@yonsei.ac.kr

affinity and molecular properties by explicitly incorporating domain knowledge through gradient-based guidance during the diffusion sampling process. Despite recent advancements, prevailing SBDD approaches optimize on-target binding affinity without penalizing off-target interactions, thereby leaving **selectivity**—the affinity gap between the intended target and off-targets—entirely uncontrolled. Pharmacologically, such an omission is fatal: promiscuous binding to unintended proteins is a major source of toxicity and a leading contributor to clinical failure [9, 18].

Achieving molecular selectivity in SBDD is fundamentally hindered by two obstacles: **(1) Spatial Reference Frame Mismatch.** Current 3D generative models construct ligands within the on-target protein's reference frame (Figure 1), precluding simultaneous consideration of off-target proteins in distinct spatial configurations. Since noisy intermediate ligands are chemically invalid, standard docking cannot be used for off-target affinity estimation during generation [16]. **(2) Multi-Objective Divergence.** On-target and off-target proteins often exhibit substantial geometric and chemical divergence in their binding environments [2, 10]. Optimizing a ligand for a high selectivity gap requires guidance that can strategically manage the trade-off between competing interaction profiles with both proteins while modulating the molecular chemical composition and geometry.

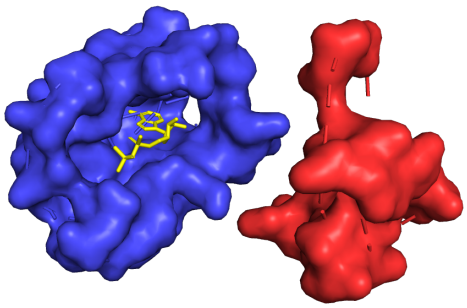

Fig. 1. Illustration of the spatial challenge in off-target consideration. The on-target protein (Left, with ligand) and the off-target protein (Right) exist in separate 3D spaces.

Motivated by these challenges, our research offers the following contributions:

- **Alignment-Free Off-Target Prediction:** To resolve obstacle **(1)**, we introduce a dual-affinity predictor framework with distinct architectures for on-target and off-target estimation. Our off-target affinity predictor utilizes a cross-attention mechanism over independent protein and ligand embeddings. This approach bypasses the need for explicit 3D spatial alignment or docking, enabling the model to estimate off-target affinity directly during the diffusion generation process.
- **Dual-Affinity Guidance for Selectivity:** To tackle obstacle **(2)**, we inject on-target attractive and off-target repulsive gradients into the denoising process, steering both atomic types and coordinates toward selective molecules. Additionally, a temporal schedule removes off-target guidance in later denoising steps, preventing residual conflicts during fine-grained molecular refinement.

2 Methodology

2.1 Problem Statement

Structure-Based Drug Design. We represent a protein pocket as a set of atoms $\mathcal{P} = \{(\mathbf{x}_i^P, \mathbf{z}_i^P)\}_{i=1}^{N_P}$, where N_P denotes the number of protein atoms. Each atom is defined by its 3D Cartesian coordinates $\mathbf{x}_i^P \in \mathbb{R}^3$ and a concatenated feature vector $\mathbf{z}_i^P \in \mathbb{R}^{d_P}$ (e.g., element type, amino acid type). A ligand molecule is defined as $\mathcal{M} = \{(\mathbf{x}_j^M, \mathbf{z}_j^M)\}_{j=1}^{N_M}$, where N_M is the number of ligand atoms. Here, $\mathbf{x}_j^M \in \mathbb{R}^3$ represents the 3D coordinates and $\mathbf{z}_j^M \in \{0,1\}^K$ represents the one-hot encoded atom type over K possible elements. In matrix notation, the ligand is denoted as $\mathcal{M} = [\mathbf{X}^M, \mathbf{Z}^M]$, where $\mathbf{X}^M \in \mathbb{R}^{N_M \times 3}$ and $\mathbf{Z}^M \in \mathbb{R}^{N_M \times K}$ represent the concatenated coordinate and feature matrices, respectively.

Problem of Interest. Given an on-target protein $\mathcal{P}^{\text{on}}$ and an off-target protein $\mathcal{P}^{\text{off}}$, our primary objective is to generate a ligand $\mathcal{M}^*$ that selectively binds the target. Formally, we aim to maximize the selectivity margin:

$$\mathcal{M}^* = \arg\max_{\mathcal{M}} \left[f(\mathcal{M}, \mathcal{P}^{\text{off}}) - f(\mathcal{M}, \mathcal{P}^{\text{on}}) \right] \tag{1}$$

where $f(\mathcal{M}, \mathcal{P})$ denotes a docking score (e.g., AutoDock Vina [19]) between $\mathcal{M}$ and $\mathcal{P}$, with more negative values indicating stronger binding.

2.2 Diffusion Framework for Molecular Generation

Diffusion models for SBDD consist of forward and reverse processes [1].

The forward process progressively adds noise to a ligand molecule $\mathcal{M}_0 = [\mathbf{X}_0^M, \mathbf{Z}_0^M]$ over T timesteps while keeping the protein $\mathcal{P}$ fixed. For atom coordinates, Gaussian noise is applied:

$$q(\mathbf{X}_t^M | \mathbf{X}_0^M) = \mathcal{N}(\mathbf{X}_t^M; \sqrt{\bar{\alpha}_t}\mathbf{X}_0^M, (1 - \bar{\alpha}_t)\mathbf{I}) \tag{2}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t}(1 - \beta_s)$ with a gradually increasing variance schedule β_t. For atom types, uniform categorical noise is applied:

$$q(\mathbf{Z}_t^M | \mathbf{Z}_0^M) = \mathcal{C}\left(\mathbf{Z}_t^M \mid \bar{\alpha}_t \mathbf{Z}_0^M + \frac{1 - \bar{\alpha}_t}{K}\mathbf{1}_K\right) \tag{3}$$

where $\mathcal{C}$ denotes a categorical distribution over K atom types and $\mathbf{1}_K$ represents a vector of ones.

The reverse process learns to iteratively denoise $\mathcal{M}_T$ back to $\mathcal{M}_0$ through a neural network ϕ_θ parameterized by θ. At each timestep t, the network predicts the clean molecule:

$$[\hat{\mathbf{X}}_0^M, \hat{\mathbf{Z}}_0^M] = \phi_\theta(\mathcal{M}_t, t, \mathcal{P}) \tag{4}$$

where $\hat{\mathbf{X}}_0^M$ and $\hat{\mathbf{Z}}_0^M$ are the predicted clean atom coordinates and types. This prediction parametrizes the posterior distribution used to sample $\mathcal{M}_{t-1}$ during inference (Section 2.5).

Multi-Task Training Objective. Following recent guidance-based approaches [16], we adopt a hybrid loss function that jointly optimizes molecule generation and binding affinity prediction:

$$\mathcal{L}_{\text{total}} = \lambda_x \mathcal{L}_{\text{coord}} + \lambda_z \mathcal{L}_{\text{type}} + \lambda_v \mathcal{L}_{\text{affinity}} \tag{5}$$

where $\lambda_x, \lambda_z, \lambda_v$ are balancing hyperparameters.

Coordinate Loss employs the mean squared error between the predicted and ground-truth atom positions at step $t = 0$, facilitating direct reconstruction:

$$\mathcal{L}_{\text{coord}} = \mathbb{E}_{t,\mathcal{M}_0,\mathcal{M}_t,\mathcal{P}} \left[\|\mathbf{X}_0^M - \hat{\mathbf{X}}_0^M\|^2 \right] \tag{6}$$

Atom Type Loss minimizes the KL divergence between the true categorical posterior $q(\mathbf{Z}_{t-1}^M | \mathbf{Z}_0^M, \mathbf{Z}_t^M)$ and the predicted posterior $p_\theta(\mathbf{Z}_{t-1}^M | \hat{\mathbf{Z}}_0^M, \mathbf{Z}_t^M)$:

$$\mathcal{L}_{\text{type}} = \mathbb{E}_{t,\mathcal{M}_0,\mathcal{M}_t,\mathcal{P}} \left[\text{KL} \left(q(\mathbf{Z}_{t-1}^M | \mathbf{Z}_0^M, \mathbf{Z}_t^M) \,\|\, p_\theta(\mathbf{Z}_{t-1}^M | \hat{\mathbf{Z}}_0^M, \mathbf{Z}_t^M) \right) \right] \tag{7}$$

Binding Affinity Loss trains two predictor heads: a complex-graph head g_ψ^{on} for on-target affinity and a cross-attention head g_ψ^{off} for off-target affinity:

$$\mathcal{L}_{\text{affinity}} = \mathbb{E}_{t,\mathcal{M}_0,\mathcal{M}_t,\mathcal{P}} \left[\|v - \hat{v}_{\text{on}}\|^2 + \|v - \hat{v}_{\text{off}}\|^2 \right] \tag{8}$$

where $v \in [0, 1]$ is the normalized ground-truth affinity, with higher values indicating stronger binding. During training, both heads receive the same protein $\mathcal{P}$ and noisy ligand $\mathcal{M}_t$ but differ in input processing: $\hat{v}_{\text{on}} = g_\psi^{\text{on}}(\mathcal{M}_t, t, \mathcal{P})$ operates on the complex graph, while $\hat{v}_{\text{off}} = g_\psi^{\text{off}}(\mathcal{M}_t, t, \mathcal{P})$ uses independent embeddings.

2.3 Generative Model Architecture

Our denoising network consists of RefineNet, an SE(3)-equivariant backbone based on EGNN [17]. The network processes the protein–ligand complex as a graph, where nodes represent atoms and edges connect spatially proximate atoms. Following the architecture of TargetDiff [7] and KGDiff [16], RefineNet updates node features $\mathbf{h}_i^{(\ell)} \in \mathbb{R}^{d_h}$ (SE(3)-invariant) and coordinates $\mathbf{x}_i^{(\ell)} \in \mathbb{R}^3$ (SE(3)-equivariant) at layer ℓ using a residual framework:

$$\mathbf{m}_{ij}^{(\ell)} = \phi_h \left(\mathbf{h}_i^{(\ell)}, \mathbf{h}_j^{(\ell)}, \|\mathbf{x}_i^{(\ell)} - \mathbf{x}_j^{(\ell)}\|^2, \mathbf{e}_{ij} \right) \tag{9}$$

$$\mathbf{x}_i^{(\ell+1)} = \mathbf{x}_i^{(\ell)} + \sum_{j \in \mathcal{N}(i)} (\mathbf{x}_i^{(\ell)} - \mathbf{x}_j^{(\ell)}) \cdot \phi_x \left(\mathbf{m}_{ij}^{(\ell)} \right) \tag{10}$$

$$\mathbf{h}_i^{(\ell+1)} = \mathbf{h}_i^{(\ell)} + \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}^{(\ell)} \tag{11}$$

where $\mathbf{m}_{ij}^{(\ell)}$ represents the interaction message from atom j to atom i, $\mathcal{N}(i)$ denotes the set of neighboring atoms, and $\mathbf{e}_{ij}$ represents edge features. Here, ϕ_h

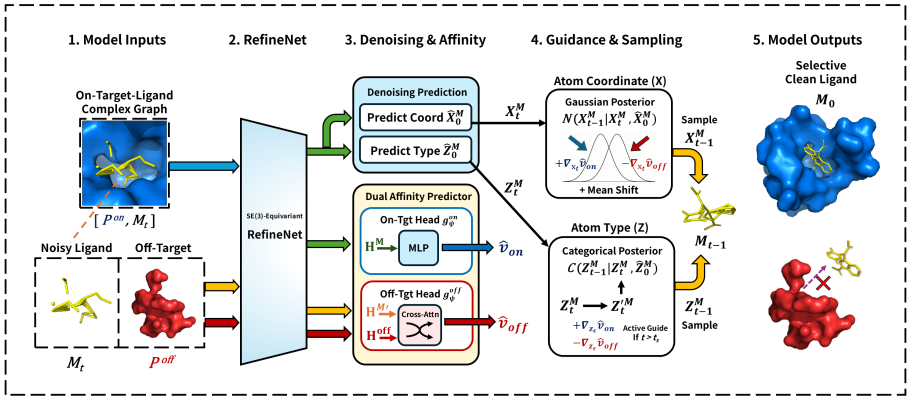

Fig. 2. Overview of **TheSelective**. The model takes an on-target–ligand complex, noisy ligand, and an off-target protein as separate inputs, encodes them via an SE(3)-equivariant RefineNet, and performs dual affinity-guided reverse diffusion to generate molecules with high on-target binding and low off-target binding.

is an MLP that computes the message embeddings, and ϕ_x outputs a scalar weight to modulate radial updates. After L layers, RefineNet produces the predicted clean ligand coordinates $\hat{\mathbf{X}}_0^M$ and refined embeddings for both protein atoms $\mathbf{H}^P \in \mathbb{R}^{N_P \times d_h}$ and ligand atoms $\mathbf{H}^M \in \mathbb{R}^{N_M \times d_h}$. The ligand embeddings $\mathbf{H}^M$ are passed to a prediction head to obtain clean atom types: $\hat{\mathbf{Z}}_0^M = \mathrm{softmax}(\mathrm{MLP}_Z(\mathbf{H}^M))$.

2.4 Asymmetric Dual Affinity Predictors

As introduced in Section 1, the ligand $\mathcal{M}_t$ is constructed within the on-target protein's coordinate system, enabling direct 3D interaction modeling for the on-target complex $[\mathcal{P}^{\mathrm{on}}, \mathcal{M}_t]$. However, the off-target protein $\mathcal{P}^{\mathrm{off}}$ resides in a separate spatial frame, rendering explicit 3D interaction modeling infeasible. To address this, we design two distinct affinity predictors: the complex-graph head g_ψ^{on} for on-target affinity and the cross-attention head g_ψ^{off} for off-target affinity.

On-Target Affinity Predictor g_ψ^{on}. The complex graph $[\mathcal{P}^{\mathrm{on}}, \mathcal{M}_t]$ is jointly processed by RefineNet, yielding ligand atom embeddings $\mathbf{H}^M$ that encode explicit 3D protein-ligand interactions. Following KGDiff [16], the on-target affinity is predicted via a per-atom MLP and mean aggregation:

$$\hat{v}_{\mathrm{on}} = \frac{1}{N_M} \sum_{j=1}^{N_M} \sigma\big(\mathrm{MLP}_{\mathrm{on}}(\mathbf{h}_j^M)\big) \tag{12}$$

where $\sigma(\cdot)$ is the sigmoid function.

Off-Target Affinity Predictor g_ψ^{off}. The ligand and off-target protein are independently processed by RefineNet, yielding interaction-free embeddings $\mathbf{H}^{M'} \in \mathbb{R}^{N_M \times d_h}$ and $\mathbf{H}^{\text{off}} \in \mathbb{R}^{N_{\text{off}} \times d_h}$, respectively. These are then processed via **bidirectional cross-attention** to predict off-target affinity.

Protein-to-Ligand Attention. Each off-target protein atom queries all ligand atoms to identify which protein regions are engaged by the ligand. The protein embeddings serve as queries ($\mathbf{Q}_1$), while the ligand embeddings serve as keys and values ($\mathbf{K}_1, \mathbf{V}_1$), and $\hat{v}_{\text{off}}^{(P)}$ is the mean of per-atom affinity predictions from the updated off-target protein embeddings:

$$\mathbf{Q}_1 = \mathbf{H}^{\text{off}}, \quad \mathbf{K}_1 = \mathbf{V}_1 = \mathbf{H}^{M'} \tag{13}$$

$$\tilde{\mathbf{H}}^{\text{off}} = \text{softmax}\left(\frac{\mathbf{Q}_1 \mathbf{K}_1^\top}{\sqrt{d_h}}\right)\mathbf{V}_1 \in \mathbb{R}^{N_{\text{off}} \times d_h} \tag{14}$$

$$\hat{v}_{\text{off}}^{(P)} = \frac{1}{N_{\text{off}}} \sum_{i=1}^{N_{\text{off}}} \sigma\Big(\text{MLP}_{\text{off}}^{(P)}(\tilde{\mathbf{h}}_i^{\text{off}})\Big) \tag{15}$$

Ligand-to-Protein Attention. Each ligand atom queries all off-target protein atoms to reveal which ligand atoms contribute to off-target binding. The ligand embeddings serve as queries ($\mathbf{Q}_2$), while the protein embeddings serve as keys and values ($\mathbf{K}_2, \mathbf{V}_2$), and $\hat{v}_{\text{off}}^{(M)}$ is the mean of per-atom affinity predictions from the updated ligand embeddings:

$$\mathbf{Q}_2 = \mathbf{H}^{M'}, \quad \mathbf{K}_2 = \mathbf{V}_2 = \mathbf{H}^{\text{off}} \tag{16}$$

$$\tilde{\mathbf{H}}^{M'} = \text{softmax}\left(\frac{\mathbf{Q}_2 \mathbf{K}_2^\top}{\sqrt{d_h}}\right)\mathbf{V}_2 \in \mathbb{R}^{N_M \times d_h} \tag{17}$$

$$\hat{v}_{\text{off}}^{(M)} = \frac{1}{N_M} \sum_{j=1}^{N_M} \sigma\Big(\text{MLP}_{\text{off}}^{(M)}(\tilde{\mathbf{h}}_j^{M'})\Big) \tag{18}$$

Aggregation. The final off-target affinity averages both perspectives:

$$\hat{v}_{\text{off}} = \frac{1}{2}\left(\hat{v}_{\text{off}}^{(P)} + \hat{v}_{\text{off}}^{(M)}\right) \tag{19}$$

2.5 Scheduled Dual-Affinity Guided Generation

Diffusion models exhibit a coarse-to-fine dynamic: early steps establish the global molecular scaffold, while later steps refine local geometric details. Applying off-target repulsive gradients during late-stage refinement risks disrupting the delicate on-target binding geometry. We therefore adopt a **scheduled guidance strategy** over the full $T = 1000$ denoising trajectory ($t : 1000 \to 0$) with a transition timestep $t_s = 500$: for $t > t_s$, both on-target and off-target signals guide generation; for $t \leq t_s$, only on-target guidance is retained.

Atom Coordinate Guidance. We apply coordinate guidance by shifting the posterior mean of the transition kernel and sampling the previous coordinates $\mathbf{X}_{t-1}^M$ from the modified Gaussian distribution $\mathcal{N}$:

$$\mathbf{X}_{t-1}^M \sim \mathcal{N}\Big(\tilde{\mu}_t(\mathbf{X}_t^M, \hat{\mathbf{X}}_0^M) + s\tilde{\beta}_t\Big(w_{\text{on}} \cdot \nabla_{\mathbf{X}_t^M} \hat{v}_{\text{on}} - \mathbb{1}_{[t>t_s]} \cdot w_{\text{off}} \cdot \nabla_{\mathbf{X}_t^M} \hat{v}_{\text{off}}\Big),\ \tilde{\beta}_t \mathbf{I}\Big) \tag{20}$$

where s is the coordinate gradient scale, $w_{\text{on/off}}$ are guidance strengths, and $\tilde{\mu}_t(\mathbf{X}_t^M, \hat{\mathbf{X}}_0^M) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{X}}_0^M + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{X}_t^M$ is the posterior mean computed from the predicted clean coordinates $\hat{\mathbf{X}}_0^M$ (Section 2.3), with posterior variance $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$.

Atom Type Guidance. At timestep t, the one-hot encoded atom types $\mathbf{Z}_t^M \in \{0,1\}^{N_M \times K}$ are relaxed with a small constant $\epsilon > 0$ to enable gradient computation on discrete variables:

$$\mathbf{Z}_\epsilon = \mathbf{Z}_t^M + \epsilon \cdot \mathbf{1} \tag{21}$$

where $\mathbf{1} \in \mathbb{R}^{N_M \times K}$ is an all-ones matrix. The guided atom types are computed by injecting the affinity gradients to the relaxed input $\mathbf{Z}_\epsilon$:

$$\log \mathbf{Z}_t'^M = \log \mathbf{Z}_\epsilon + r\big(w_{\text{on}} \cdot \nabla_{\mathbf{Z}_\epsilon} \hat{v}_{\text{on}} - \mathbb{1}_{[t>t_s]} \cdot w_{\text{off}} \cdot \nabla_{\mathbf{Z}_\epsilon} \hat{v}_{\text{off}}\big) \tag{22}$$

where r is the atom type gradient scale. The guided $\mathbf{Z}_t'^M$ is fed into the categorical posterior $\mathcal{C}$ to sample $\mathbf{Z}_{t-1}^M$:

$$\mathbf{Z}_{t-1}^M \sim \mathcal{C}\Big(\tilde{c}_t(\mathbf{Z}_t'^M, \hat{\mathbf{Z}}_0^M)\Big) \tag{23}$$

where $\tilde{c}_t$ computes the categorical posterior probabilities [7] and $\hat{\mathbf{Z}}_0^M$ is the predicted clean atom types (Section 2.3).

3 Experiments and Results

3.1 Experimental Setup

Dataset. We employed the CrossDocked2020 dataset [6] for both training and evaluation. Following standard practices [7,15], the dataset was filtered (RMSD < 1Å; sequence identity $< 30\%$ between splits), resulting in approximately 100,000 training pairs and 100 test proteins.

To systematically evaluate selectivity across different structural similarity regimes, we constructed two test sets based on pairwise **TM-scores** [20], defined as:

$$\text{TM-score} = \max\left[\frac{1}{L_{\text{target}}}\sum_{i=1}^{L_{\text{ali}}}\frac{1}{1+\left(\frac{d_i}{d_0}\right)^2}\right]$$

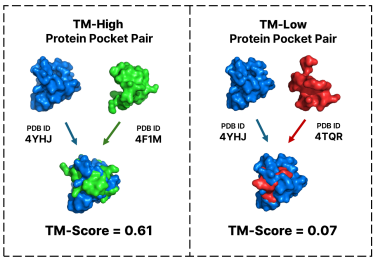

Fig. 3. Examples of TM-Score.

where L_{target} is the target protein length, L_{ali} is the number of aligned residue pairs, d_i is the distance between the i-th pair, and d_0 is a normalization threshold. The TM-score ranges in $(0, 1]$, with 1 indicating perfect structural match. For each test pocket, the **TM-High** pairs it with the most structurally similar protein, while the **TM-Low** pairs it with a geometrically distinct protein.

Baselines and Evaluation Metrics. We compare against several SBDD baselines: TargetDiff [7], KGDiff [16], and BInD [12]. For each test pair, each model generates a fixed number of ligand candidates, and all reported metrics follow the same aggregation: **Avg.** denotes the mean of per-pair average scores, and **Med.** denotes the mean of per-pair median scores. Generated ligands are assessed in terms of **binding affinity** and **molecular properties**. Following previous work [7, 13, 16], we use AutoDock Vina [19] to compute on-target and off-target docking scores; **Selectivity** is defined as the difference between them. For molecular properties, we report **QED** [3] (drug-likeness) and **SA** [5] (synthetic accessibility), both normalized to $[0, 1]$, with higher being better. **Success** denotes the proportion of molecules validly reconstructed by Open Babel [14] that achieve successful docking outcomes for both proteins.

3.2 Performance on TM-Score Datasets

Tables 1 and 2 summarize the overall performance on TM-High and TM-Low protein pairs, respectively, with the best results highlighted in **bold**.

Table 1. Evaluation Results of TM-High Protein Pairs

Metrics / Models	On-Dock(↓)		Off-Dock(↑)		Selectivity(↑)		QED(↑)		SA(↑)		Success(↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Value
Reference	-7.400	-7.179	-7.168	-7.056	0.231	0.083	0.482	0.469	0.736	0.745	98%
BInD [12]	-7.510	-7.569	**-7.391**	-7.422	0.120	0.062	0.505	0.503	**0.658**	**0.661**	88.1%
TargetDiff [7]	-7.583	-7.577	-7.405	**-7.391**	0.178	0.165	0.469	0.466	0.585	0.591	**91.0%**
KGDiff [16]	-9.290	-9.309	-8.446	-8.466	0.844	0.727	**0.527**	**0.537**	0.548	0.550	85.6%
TheSelective	**-9.969**	**-9.958**	-8.994	-9.001	**0.975**	**0.923**	0.495	0.500	0.534	0.535	50.2%

Table 2. Evaluation Results of TM-Low Protein Pairs

Metrics / Models	On-Dock(↓)		Off-Dock(↑)		Selectivity(↑)		QED(↑)		SA(↑)		Success(↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Value
Reference	-7.451	-7.293	-5.414	-5.424	2.037	1.869	0.475	0.469	0.728	0.740	97%
BInD [12]	-7.536	-7.589	-5.608	-5.643	1.928	1.934	0.502	0.500	**0.654**	**0.656**	89.1%
TargetDiff [7]	-7.566	-7.552	**-5.567**	**-5.564**	1.999	1.993	0.467	0.464	0.583	0.591	**91.9%**
KGDiff [16]	-9.343	-9.366	-6.345	-6.393	2.998	2.980	**0.528**	**0.538**	0.546	0.549	85.6%
TheSelective	**-9.954**	**-9.909**	-6.591	-6.588	**3.363**	**3.259**	0.511	0.514	0.557	0.561	56.9%

For TM-High pairs (Table 1), TheSelective achieves the highest selectivity ($\Delta_{\text{avg}} = 0.975$, $\Delta_{\text{med}} = 0.923$ kcal/mol), outperforming the strongest baseline KGDiff ($\Delta_{\text{avg}} = 0.844$, $\Delta_{\text{med}} = 0.727$) by 15.5% and 27.0% improvements in average and median, respectively. TheSelective also attains the best on-target docking scores (-9.969 avg., -9.958 med.), indicating that the selectivity gains do not compromise on-target binding strength.

For TM-Low pairs (Table 2), TheSelective again achieves the best selectivity ($\Delta_{\text{avg}} = 3.363$, $\Delta_{\text{med}} = 3.259$), surpassing KGDiff ($\Delta_{\text{avg}} = 2.998$, $\Delta_{\text{med}} = 2.980$) with improvements of 12.2% and 9.4%, while maintaining the strongest on-target docking (-9.954 avg., -9.909 med.). The consistent improvements across both settings demonstrate that our model reliably generates selective ligands regardless of the structural relationship between target and off-target proteins.

Molecular Quality Assessment. Across both settings, TheSelective maintains comparable QED and SA scores to the baselines. As noted in prior work [7, 16], these metrics serve as rough filters rather than primary optimization targets, and our results fall within an acceptable range. We observe a reduction in Success Rate, which reflects the inherent tension of jointly optimizing for on-target potency and off-target avoidance. Integrating validity-aware constraints into the guidance process is a promising direction for future work.

3.3 Ablation Analysis of Guidance Contributions

Table 3 presents ablation studies to validate the contribution of each component in our guidance mechanism. We compare five configurations: **No Guidance** (baseline diffusion without guidance), **On-Tgt Guide** (maximizing on-target affinity only), **Off-Tgt Guide** (minimizing off-target affinity only), **Dual Guide** (utilizing both signals throughout all timesteps), and **Dual Guide, Scheduled** (our complete framework with scheduling).

Table 3. Experimental Results of Guidance Effects

Data	TM-High						TM-Low					
Metrics	On-Dock(↓)		Off-Dock(↑)		Selectivity(↑)		On-Dock(↓)		Off-Dock(↑)		Selectivity(↑)	
Methods	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.
No Guide	-7.894	-7.875	**-7.729**	**-7.724**	0.165	0.112	-7.892	-7.864	**-5.771**	**-5.769**	2.121	2.064
On-Tgt Guide	-9.204	-9.098	-8.568	-8.621	0.636	0.536	-9.193	-9.066	-6.349	-6.340	2.843	2.756
Off-Tgt Guide	-7.909	-7.856	-7.873	-7.806	0.036	0.047	-8.095	-7.998	-5.844	-5.774	2.251	2.216
Dual Guide	-9.302	-9.334	-8.673	-8.672	0.628	0.588	-9.306	-9.252	-6.353	-6.367	2.953	2.890
Dual Guide, Scheduled	**-9.969**	**-9.958**	-8.994	-9.001	**0.975**	**0.923**	**-9.954**	**-9.909**	-6.591	-6.588	**3.363**	**3.259**

We observe that on-target guidance alone improves on-target affinity but yields limited selectivity, while off-target guidance alone provides insufficient

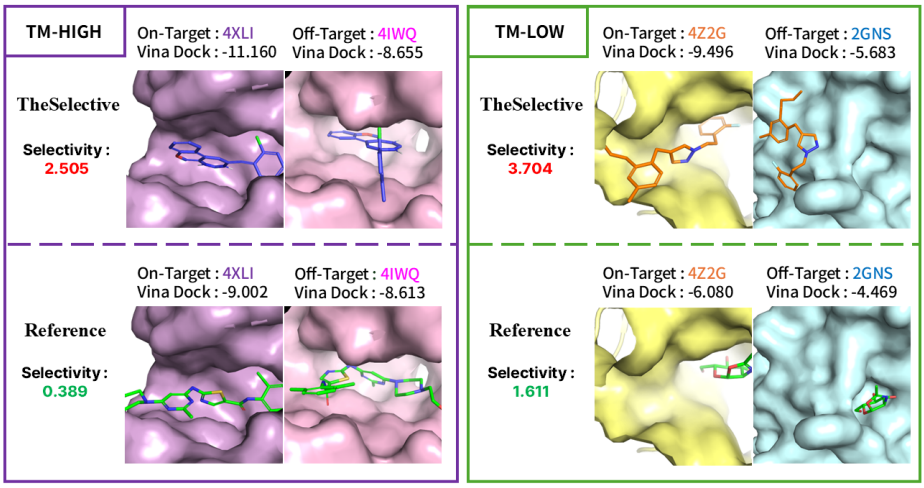

Fig. 4. Case studies comparing TheSelective-generated molecules (top) with reference ligands (bottom). Left: High TM-score pair (4XLI/4IWQ). Right: Low TM-score pair (4Z2G/2GNS).

direction for strong on-target binding. The dual guidance configuration outperforms either single-guidance variant, confirming that complementary modeling of both binding signals is essential for selective generation. Furthermore, applying temporal scheduling to the dual guidance yields the highest selectivity across both TM-score regimes, as transitioning to on-target-only guidance at later timesteps allows the model to first establish selectivity-aware scaffolds and then refine on-target binding without continued off-target perturbation.

3.4 Case Studies

To validate the practical applicability of our framework, we analyze generated molecules against two representative protein pairs with distinct structural similarity profiles. Figure 4 illustrates the binding poses and selectivity metrics compared to reference ligands.

High TM-Score Case: We examined Abelson tyrosine-protein kinase 2 (PDB ID: 4XLI, on-target) paired with the structurally similar TANK-binding kinase 1 (PDB ID: 4IWQ, off-target). The reference ligand shows poor selectivity ($\Delta = 0.389$ kcal/mol), with comparable affinities for both targets. In contrast, TheSelective generates a ligand that substantially amplifies selectivity ($\Delta = 2.505$ kcal/mol), primarily driven by a marked enhancement in on-target affinity (-11.160 kcal/mol). This result suggests that our guidance mechanism exploits subtle differences in the ATP-binding site architectures of homologous kinases to prioritize on-target interactions.

Low TM-Score Case: We compared chitinase B from *Serratia marcescens* (PDB ID: 4Z2G, on-target) with the structurally distinct phospholipase A_2 from *Daboia russelii* (PDB ID: 2GNS, off-target). While the reference ligand exhibits moderate selectivity ($\Delta = 1.611$ kcal/mol), TheSelective achieves a superior selectivity margin ($\Delta = 3.704$ kcal/mol). The generated molecule optimizes geometric and chemical complementarity within the 4Z2G pocket (-9.496 kcal/mol) through favorable interactions with the deep substrate-binding cleft of chitinase B, while maintaining a significant affinity gap against the off-target (-5.683 kcal/mol). These results indicate that, even for structurally dissimilar protein pairs, our guidance strategy effectively leverages the divergent binding pocket geometries to produce ligands with enhanced selectivity.

4 Conclusion

In this work, we presented **TheSelective**, a diffusion-based SBDD framework that explicitly addresses molecular selectivity by incorporating both on-target and off-target binding affinity signals into the generative process through a scheduled dual-guidance mechanism. Empirical evaluations on the CrossDocked2020 dataset [6] confirm that TheSelective consistently achieves the highest selectivity across both TM-High and TM-Low protein pair regimes, demonstrating its effectiveness regardless of the structural relationship between on-target and off-target proteins. Ablation studies further validate that the combination of dual guidance with temporal scheduling is essential for achieving high selectivity. Case studies on representative protein pairs show that TheSelective can generate ligands with substantially improved selectivity margins. Despite these promising results, we observe that the dual-objective guidance leads to lower success rates in molecular reconstruction. Future work will explore multi-objective strategies that jointly enhance selectivity, structural validity, QED, and SA. Nonetheless, our framework establishes a practical foundation for incorporating selectivity as an explicit design objective in structure-based molecular generation, taking a step toward therapeutic candidates that balance potency with target discrimination.

Acknowledgments. This research was supported by the National Research Foundation (NRF) funded by the Korean government (MSIT) (No. RS-2023-00229822).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alakhdar, A., Poczos, B., Washburn, N.: Diffusion models in de novo drug design. Journal of Chemical Information and Modeling **64**(19), 7238–7256 (2024)
2. Albanese, S.K., Chodera, J.D., Volkamer, A., Keng, S., Abel, R., Wang, L.: Is structure-based drug design ready for selectivity optimization? Journal of chemical information and modeling **60**(12), 6211–6227 (2020)

3. Bickerton, G.R., Paolini, G.V., Besnard, J., Muresan, S., Hopkins, A.L.: Quantifying the chemical beauty of drugs. Nature chemistry **4**(2), 90–98 (2012)
4. Dorna, V., Subhalingam, D., Kolluru, K., Tuli, S., Singh, M., Singal, S., Krishnan, N., Ranu, S.: Tagmol: Target-aware gradient-guided molecule generation. arXiv preprint arXiv:2406.01650 (2024)
5. Ertl, P., Schuffenhauer, A.: Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. Journal of cheminformatics **1**(1), 8 (2009)
6. Francoeur, P.G., Masuda, T., Sunseri, J., Jia, A., Iovanisci, R.B., Snyder, I., Koes, D.R.: Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. Journal of chemical information and modeling **60**(9), 4200–4215 (2020)
7. Guan, J., Qian, W.W., Peng, X., Su, Y., Peng, J., Ma, J.: 3d equivariant diffusion for target-aware molecule generation and affinity prediction. arXiv preprint arXiv:2303.03543 (2023)
8. Guan, J., Zhou, X., Yang, Y., Bao, Y., Peng, J., Ma, J., Liu, Q., Wang, L., Gu, Q.: Decompdiff: diffusion models with decomposed priors for structure-based drug design. arXiv preprint arXiv:2403.07902 (2024)
9. Harrison, R.K.: Phase ii and phase iii failures: 2013–2015. Nature reviews Drug discovery **15**(12), 817–818 (2016)
10. Huggins, D.J., Sherman, W., Tidor, B.: Rational approaches to improving selectivity in drug design. Journal of medicinal chemistry **55**(4), 1424–1444 (2012)
11. Jian, Y., Wu, C., Reidenbach, D., Krishnapriyan, A.S.: General binding affinity guidance for diffusion models in structure-based drug design. arXiv preprint arXiv:2406.16821 (2024)
12. Lee, J., Zhung, W., Seo, J., Kim, W.Y.: Bind: Bond and interaction-generating diffusion model for multi-objective structure-based drug design. Advanced Science **12**(35), e02702 (2025)
13. Luo, S., Guan, J., Ma, J., Peng, J.: A 3d generative model for structure-based drug design. Advances in Neural Information Processing Systems **34**, 6229–6239 (2021)
14. O'Boyle, N.M., Banck, M.S., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R.: Open babel: An open chemical toolbox. Journal of Cheminformatics **3**, 33 – 33 (2011), https://api.semanticscholar.org/CorpusID:4857085
15. Peng, X., Luo, S., Guan, J., Xie, Q., Peng, J., Ma, J.: Pocket2mol: Efficient molecular sampling based on 3d protein pockets. In: International conference on machine learning. pp. 17644–17655. PMLR (2022)
16. Qian, H., Huang, W., Tu, S., Xu, L.: Kgdiff: towards explainable target-aware molecule generation with knowledge guidance. Briefings in Bioinformatics **25**(1), bbad435 (2024)
17. Satorras, V.G., Hoogeboom, E., Welling, M.: E (n) equivariant graph neural networks. In: International conference on machine learning. pp. 9323–9332. PMLR (2021)
18. Shityakov, S., Förster, C.: In silico predictive model to determine vector-mediated transport properties for the blood–brain barrier choline transporter. Advances and Applications in Bioinformatics and Chemistry pp. 23–36 (2014)
19. Trott, O., Olson, A.J.: Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry **31**(2), 455–461 (2010)
20. Zhang, Y., Skolnick, J.: Tm-align: a protein structure alignment algorithm based on the tm-score. Nucleic acids research **33**(7), 2302–2309 (2005)