TD3B: TRANSITION-DIRECTED DISCRETE DIFFUSION FOR ALLOSTERIC BINDER GENERATION

Hanqun Cao,[1,*] **Aastha Pal,**[2,*] **Sophia Tang,**[3] **Yinuo Zhang,**[3,4] **Jingjie Zhang,**[1] **Pheng-Ann Heng**,[1] **Pranam Chatterjee**[2,3,†]

[1]Department of Computer Science and Engineering, The Chinese University of Hong Kong
[2]Department of Bioengineering, University of Pennsylvania
[3]Department of Computer and Information Science, University of Pennsylvania
[4]Centre for Computational Biology, Duke-NUS Medical School

[*]These authors contributed equally
[†]Corresponding author: pranam@seas.upenn.edu

ABSTRACT

Protein function is often controlled by ligands that bias the direction of state transitions, such as agonists and antagonists, rather than stabilizing a single conformation. This is especially important for clinically relevant G protein-coupled receptors (GPCRs), where therapeutic efficacy depends on functional directionality. Structure-based design methods optimize binding to static conformations and cannot represent non-reversible, directional effects or systematically distinguish agonist from antagonist behavior. To address this gap, we introduce **T**ransition-**D**irected **D**iscrete **D**iffusion for allosteric **B**inder design (**TD3B**), a sequence-based generative framework that designs binders with specified agonist or antagonist behavior via a directional transition control objective. TD3B combines a target-aware Direction Oracle, a soft binding-affinity gate, and amortized fine-tuning of a pre-trained discrete diffusion model, enabling targeted agonist and antagonist generation decoupled from binding affinity and unattainable by equilibrium-based or inference-only guidance baselines.

1 INTRODUCTION

Protein allostery governs regulation and control across signaling, transport, and transcriptional systems, yet its functional effects are fundamentally dynamic (Motlagh et al., 2014; Weikl & Paul, 2014). In many biologically relevant settings, most notably for agonists and antagonists, function depends not on stabilizing a single conformation, but on biasing the direction of transitions between macrostates such as activation and deactivation (Henzler-Wildman & Kern, 2007; Cao et al., 2021). Despite this, most contemporary binder design algorithms, such as RFdiffusion (Watson et al., 2023), BindCraft (Pacesa et al., 2025), BoltzGen (Stark et al., 2025), and GPCR-specialized extensions like RareFoldGPCR (Li et al., 2025), treat proteins as fixed objects and frame design around stabilizing a target structure or interface. While these methods can generate binders or even active agonists, they remain intrinsically tied to equilibrium structural priors and lack a mechanism to represent or control transition asymmetry. As a result, they cannot distinguish or systematically design agonist versus antagonist behavior, since static structures alone do not encode non-reversible, directional effects. Consequently, binders whose function arises from reshaping kinetic pathways rather than stabilizing endpoints lie outside the representational scope of structure-centric design approaches.

Recent advances in discrete generative modeling have produced high-capacity peptide language models that capture the structure of peptide sequence space independent of downstream objectives (Tang et al., 2025a;c; Chen et al., 2025b;a; Vincoff et al., 2025; Tang et al., 2025b). In particular, masked discrete diffusion language models (MDLMs) such as PepMDLM learn strong unconditional priors over valid peptide sequences, decoupling sequence syntax and diversity from task-specific optimization (Tang et al., 2025a;c). Building on this foundation, guidance strategies such as PepTune and TR2-D2 introduce lightweight fine-tuning and objective-guided sampling mechanisms that

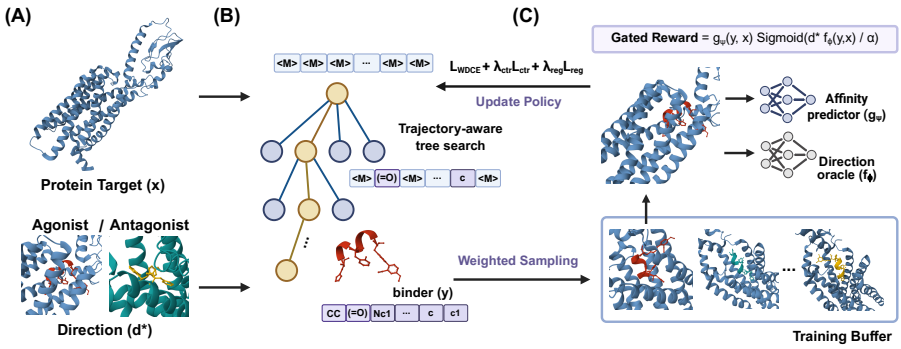

Figure 1: **Overview of the TD3B framework.** **(A)** TD3B enables flexible peptide binder design with specified directionality for diverse protein targets. **(B)** Sampling phase: TD3B performs trajectory-aware tree search conditioned on the target and desired direction, generating binder candidates through weighted sampling. **(C)** Finetuning phase: The policy model is updated using samples from the training buffer, guided by the gated reward.

bias generation toward desired properties without retraining the base model from scratch (Tang et al., 2025a;c). Our approach leverages this separation by treating directional allosteric control as a guidance objective layered on top of an existing discrete diffusion backbone, rather than as a new generative architecture. **The central question we address is:** *what form should such a guidance objective take when the desired effect is directional modulation of protein states rather than equilibrium binding?*

In this work, we introduce **T**ransition-**D**irected **D**iscrete **D**iffusion for allosteric **B**inder design (**TD3B**), a discrete generative framework that treats directional allostery as a key design objective. We model binder action through *sequence-conditioned transition operators* over protein macrostates, explicitly accommodating non-reversible, antisymmetric state changes. Our approach leverages directional supervision to bias generation toward binders that promote or suppress specific state transitions. This reframes binder design as a non-equilibrium transport problem, enabling the generation of binders that control protein function by reshaping transition directionality rather than stabilizing conformations.

Our contributions are threefold:

1. **A transition-operator formulation of directional allosteric control.** We formalize binder-mediated allostery as a sequence-conditioned transition operator over protein macrostates, making directionality and non-reversibility explicit modeling targets beyond static or equilibrium-based representations.
2. **A directionally guided generative framework for binder design.** We introduce a discrete generative modeling approach that fine-tunes an existing peptide generator using direction-only supervision, enabling targeted biasing of transition directionality.
3. **Directional binder design beyond static models.** We demonstrate that contrastive, direction-based fine-tuning produces binders that selectively bias agonistic transitions while minimally affecting reverse transitions, capturing functional behaviors.

2 Preliminaries

This section reviews the generative modeling primitives and dynamical abstractions underlying our approach. We describe masked discrete diffusion language models (MDLMs) as a general-purpose backbone for sequence generation (Sahoo et al., 2024; Austin et al., 2021; Shi et al., 2024), introduce amortized objective-guided fine-tuning as a mechanism for incorporating external objectives (Tang et al., 2025c), and define a coarse-grained dynamical representation of protein function in terms of macrostates and transition operators. These components are standard or well-established; we defer the specifics of our proposed method to Section 3.

2.1 Discrete Sequence Spaces

Let $\mathcal{A}$ denote a finite alphabet and $\mathcal{A}^L$ the space of length-L sequences. In this work, sequences correspond to binders, and the formulation applies to arbitrary sequence-based binders, with peptides serving as a concrete instantiation. We write $y \in \mathcal{A}^L$ for a clean sequence and use x to denote target protein sequences.

2.2 Masked Discrete Diffusion Language Models

Masked discrete diffusion language models (MDLMs) define generative processes over discrete sequences by progressively denoising corrupted inputs. A forward noising process $q_t(x_t \mid x)$ maps a clean sequence x to a partially corrupted version x_t at time $t \in [0, 1]$, typically by masking tokens according to a time-dependent schedule. The reverse process is parameterized by a neural network that predicts the distribution of masked tokens conditioned on the unmasked context and time.

Training proceeds by minimizing a denoising cross-entropy objective:

$$\mathcal{L}_{\text{DCE}}(\theta, \boldsymbol{x}) = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\left[\frac{1}{t}\mathbb{E}_{q_t(\tilde{\boldsymbol{x}}_t|\boldsymbol{x})}\sum_{\ell:\boldsymbol{x}_t^\ell=\boldsymbol{M}} -\log p_\theta\left(\boldsymbol{x}^\ell \mid \boldsymbol{x}_t^{\text{UM}}, t\right)_{\boldsymbol{x}^\ell}\right], \tag{1}$$

where $(x_t)_{\text{UM}}$ denotes the unmasked tokens. Once trained, an MDLM defines an unconditional distribution over valid sequences and can be sampled by iteratively denoising from a fully masked input. Importantly, the MDLM captures the combinatorial structure of sequence space independently of any downstream task.

2.3 Amortized Objective-Guided Fine-Tuning of MDLMs

Let $p_{\theta_0}(x)$ denote a pre-trained MDLM defining an unconditional distribution over sequences. Objective-guided sequence design seeks to bias this distribution toward sequences that score highly under an external objective $S(x)$, while preserving the structural prior learned by p_{θ_0}.

Amortized fine-tuning methods, such as TR2-D2 (Tang et al., 2025c), achieve this by learning a new parameterization $p_\theta(x)$ that approximates a reward-tilted target distribution:

$$p^*(x) \;\propto\; p_{\theta_0}(x)\,\exp\!\left(\frac{S(x)}{\alpha}\right), \tag{2}$$

where $\alpha > 0$ controls the strength of deviation from the pre-trained prior. This formulation defines an implicit energy-based reweighting of the base model without modifying the corruption process or denoising schedule.

Since direct sampling from p^* is intractable, amortized fine-tuning optimizes θ via a weighted denoising cross-entropy (WDCE) objective. Let $\boldsymbol{X}_{0:T}$ denote a denoising trajectory with final sequence $\boldsymbol{x} = \boldsymbol{X}_T$. The fine-tuning objective is:

$$\begin{aligned}\mathcal{L}_{\text{WDCE}}(\theta) = \mathbb{E}_{p_{\text{target}}(\boldsymbol{x})}[\mathcal{L}_{\text{DCE}}(\theta; \boldsymbol{x})] &= \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^\star}[\mathcal{L}_{\text{DCE}}(\theta; \boldsymbol{X}_T)] \\ &= \mathbb{E}_{\boldsymbol{X}_{0:T}\sim\mathbb{P}^v}\Bigg[\underbrace{\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^v}(\boldsymbol{X}_{0:T})}_{w(\boldsymbol{X}_{0:T})}\,\mathcal{L}_{\text{DCE}}(\theta; \boldsymbol{X}_T)\Bigg],\end{aligned} \tag{3}$$

where $\mathbb{P}^v$ is a proposal distribution over trajectories and $w(\boldsymbol{X}_{0:T})$ is the importance weight. Under the assumption that the trajectory distribution factorizes and the reward depends only on the final sequence, the trajectory-level importance weight decomposes as:

$$w(\boldsymbol{X}_{0:T}) \propto \exp\!\left(\frac{S(\boldsymbol{X}_T)}{\alpha}\right) \cdot \prod_t \frac{p_{\theta_0}(\boldsymbol{X}_{t-1} \mid \boldsymbol{X}_t)}{p_{\bar{\theta}}(\boldsymbol{X}_{t-1} \mid \boldsymbol{X}_t)}, \tag{4}$$

where the first term captures the reward contribution and the second term accounts for the mismatch between the pre-trained and proposal transition kernels.

This procedure internalizes objective S into the model's sampling distribution, minimizing the need for extensive inference-time search. While lightweight methods like best-of-M selection can still

refine quality, this amortized initialization is notably more efficient than guidance-only approaches. Fine-tuning updates only the denoising conditionals, leaving the diffusion and corruption schedules intact.

2.4 Protein Macrostates and Coarse-Grained State Shifts

To reason about allosteric function, we adopt a coarse-grained description of protein state shifts. Let

$$\mathcal{S} = \{s_1, \ldots, s_K\} \tag{5}$$

denote a finite set of macrostates corresponding to functionally distinct configurations, such as inactive/active or closed/open. In the absence of a binder, protein state shifts are modeled as a continuous-time Markov chain (CTMC) with generator:

$$Q_0 : \mathcal{S} \times \mathcal{S} \to \mathbb{R}, \qquad Q_0(s_i, s_i) = -\sum_{j \neq i} Q_0(s_i, s_j). \tag{6}$$

This abstraction captures state-to-state transition behavior without committing to atomistic trajectories or detailed kinetic models.

2.5 Sequence-Conditioned Transition Operators

A binder sequence $y \in \mathcal{A}^L$ may alter the transition structure of proteins. We represent this effect by a sequence-conditioned generator:

$$Q^{(y)} = Q_0 + \Delta Q^{(y)}, \tag{7}$$

where $\Delta Q^{(y)}$ denotes a sequence-dependent perturbation of transition rates. No symmetry or reversibility is assumed. In general,

$$Q^{(y)}(s_i, s_j) \neq Q^{(y)}(s_j, s_i), \tag{8}$$

and the resulting dynamics need not satisfy detailed balance:

$$\pi^{(y)}(s_i) Q^{(y)}(s_i, s_j) \neq \pi^{(y)}(s_j) Q^{(y)}(s_j, s_i), \tag{9}$$

While the stationary distribution $\pi^{(y)}$ exists for finite irreducible CTMCs, it does not imply reversibility; thus, these generators lack a scalar energy gradient representation.

2.6 Directional Asymmetry

For any ordered state pair (s_i, s_j), we define the directional asymmetry induced by a sequence y as:

$$\Delta_{ij}(y) := Q^{(y)}(s_i, s_j) - Q^{(y)}(s_j, s_i). \tag{10}$$

This quantity captures the net bias of transitions between macrostates. Directional allosteric effects correspond to consistent signs of $\Delta_{ij}(y)$ for selected transitions. We emphasize that neither the absolute magnitude of $\Delta_{ij}(y)$ nor the full generator $Q^{(y)}$ is assumed to be observable; only coarse directional information may be available in practice.

The concepts introduced define the modeling primitives throughout this paper. We next specify how these components combine into a concrete generative learning problem.

3 Problem Formulation

We formalize directional allosteric binder design as an amortized objective-guided sequence generation problem, where a pre-trained MDLM is fine-tuned to generate sequences biasing protein state transitions in a specified direction. The preliminaries are shown in Appendix 2.

3.1 Data and Directional Supervision

We assume access to a dataset:

$$\mathcal{D} = \{(x^{(n)}, y^{(n)}, a^{(n)})\}_{n=1}^{N}, \tag{11}$$

where $x^{(n)}$ is the target protein sequence, $y^{(n)} \in \mathcal{A}^L$ is a binder sequence, and

$$a^{(n)} \in \{\text{full agonist},\ \text{partial agonist},\ \text{antagonist},\ \text{negative}\} \tag{12}$$

is a categorical functional action label. Negative labels indicate lack of binding and are excluded from directional supervision.

For each target protein, we adopt a two-macrostate abstraction:

$$\mathcal{S} = \{s_{\text{inactive}}, s_{\text{active}}\}. \tag{13}$$

A binder sequence y induces a sequence-conditioned generator $Q^{(y)}$ on $\mathcal{S}$, but neither $Q^{(y)}$ nor its transition rates are observed. Instead, functional labels specify the sign of the induced transition asymmetry:

$$\Delta(y) := Q^{(y)}(s_{\text{inactive}}, s_{\text{active}}) - Q^{(y)}(s_{\text{active}}, s_{\text{inactive}}). \tag{14}$$

We encode supervision using a direction label $d(y) \in \{+1, -1\}$ and a confidence weight $\kappa(y) \in [0, 1]$:

$$d(y) = \begin{cases} +1, & a(y) \in \{\text{full agonist, partial agonist}\}, \\ -1, & a(y) = \text{antagonist}, \end{cases}, \quad \kappa(y) = \begin{cases} 1, & a(y) = \text{full agonist}, \\ \kappa_{\text{part}}, & a(y) = \text{partial agonist}, \\ 1, & a(y) = \text{antagonist}, \\ 0, & a(y) = \text{negative}, \end{cases} \tag{15}$$

where $\kappa_{\text{part}} \in (0, 1)$ is a hyperparameter reflecting lower confidence in partial agonism.

3.2 Direction Oracle

We introduce a Direction Oracle $f_\phi : \mathcal{A}^L \times \mathcal{X} \to [-1, 1]$, parameterized by ϕ, that predicts the direction of transition bias. The predicted direction is defined as:

$$\hat{d}(y) = \text{sign}(f_\phi(y, x)). \tag{16}$$

Given a target protein sequence x and a peptide binder y, the representations are obtained through pre-trained encoders:

$$\mathbf{h}_x = \text{Pool}(\mathcal{E}_x(x)), \quad \mathbf{h}_y = \text{Pool}(\mathcal{E}_y(y)), \tag{17}$$

where $\mathbf{h}_x \in \mathbb{R}^d$ and $\mathbf{h}_y \in \mathbb{R}^d$ denote the pooled embeddings for the target and binder. The fused representation is computed via a gated mechanism followed by an MLP:

$$\mathbf{z} = \mathbf{g} \odot \mathbf{h}_x + (1 - \mathbf{g}) \odot \mathbf{h}_y, \quad f_\phi(y, x) = \text{MLP}(\mathbf{z}), \tag{18}$$

where $\mathbf{g} = \sigma(\mathbf{W}_g[\mathbf{h}_x; \mathbf{h}_y] + \mathbf{b}_g)$ is a learned gating vector and $\odot$ denotes element-wise multiplication.

The oracle minimizes a weighted binary classification loss:

$$\mathcal{L}_{\text{dir}}(\phi) = \mathbb{E}_{(x,y,d)\sim\mathcal{D}} \left[\kappa(y)\ \log(1 + \exp(-d \cdot f_\phi(y, x)))\right],$$

where $d \in \{-1, +1\}$ denotes the ground-truth direction and $\kappa(y)$ is a sample-dependent weight.

3.3 Contrastive Directional Representation

Let $h_\theta(y) \in \mathbb{R}^m$ denote a sequence representation extracted from the MDLM by mean-pooling the final-layer hidden states across all sequence positions. To enforce separation between directional classes in representation space, we define positive and negative index sets:

$$\mathcal{P} = \{(i, j) : d(y_i) = d(y_j),\ \kappa(y_i)\kappa(y_j) > 0\}, \tag{19}$$

$$\mathcal{N} = \{(i, j) : d(y_i) \neq d(y_j),\ \kappa(y_i)\kappa(y_j) > 0\}. \tag{20}$$

A margin-based contrastive loss is defined as:

$$\mathcal{L}_{\text{ctr}}(\theta) = \sum_{(i,j)\in\mathcal{P}} \|h_\theta(y_i) - h_\theta(y_j)\|_2^2 + \sum_{(i,j)\in\mathcal{N}} \max(0,\ m - \|h_\theta(y_i) - h_\theta(y_j)\|_2)^2, \tag{21}$$

where $m > 0$ is a margin hyperparameter. Negative (non-binding) samples are excluded from this loss.

3.4 Incorporating Target Binding Affinity via Gating

Directional allosteric control is only meaningful for sequences that actually bind to the target protein. To ensure that directional supervision is applied to plausible binders, we incorporate a pre-trained peptide-protein affinity predictor as a soft gate within the reward function.

Let

$$g_\psi(y, x) \in [0, 1] \tag{22}$$

denote a pre-trained affinity model that predicts the probability that peptide y binds target protein x.

Given a desired direction $d^\star \in \{+1, -1\}$, we define the gated reward as:

$$R(y; d^\star, x) = g_\psi(y, x) \cdot \sigma\left(\frac{d^\star \cdot f_\phi(y, x)}{\tau}\right), \tag{23}$$

where σ denotes the sigmoid function and τ is a temperature coefficient. This formulation ensures that sequences predicted not to bind contribute negligible reward regardless of their directional score, while sequences predicted to bind are ranked according to their directional effect. Crucially, binding affinity acts as a gate to filter implausible candidates, not as a quantity to be maximized—stronger binders do not necessarily induce desired state changes.

The resulting reward-tilted target distribution becomes:

$$p^\star(y \mid d^\star, x) \;\propto\; p_{\theta_0}(y)\, \exp\!\left(\frac{R(y; d^\star, x)}{\alpha}\right), \tag{24}$$

where $\alpha > 0$ controls the strength of deviation from the pre-trained prior.

Following the trajectory-level importance weighting framework from Section 2.3, the unnormalized log importance weight for a trajectory $\boldsymbol{X}_{0:T}$ with final sequence $y = \boldsymbol{X}_T$ is:

$$\log \tilde{w}(\boldsymbol{X}_{0:T}) = \frac{R(y; d^\star, x)}{\alpha} + \sum_{t=1}^{T} \sum_{\ell : \boldsymbol{X}_{t-1}^\ell \neq \boldsymbol{X}_t^\ell} \log \frac{p_{\theta_0}(\boldsymbol{X}_{t-1}^\ell \mid \boldsymbol{X}_t^{\mathrm{UM}})}{p_{\bar{\theta}}(\boldsymbol{X}_{t-1}^\ell \mid \boldsymbol{X}_t^{\mathrm{UM}})}, \tag{25}$$

where p_{θ_0} and $p_{\bar{\theta}}$ denote the pre-trained model and proposal policy. At inference, a target protein x and direction $d^\star \in \{+1, -1\}$ (agonist/antagonist) are provided. These inputs condition the reward function alone, keeping the generative backbone target-agnostic.

3.5 Amortized Fine-Tuning Objective

Direct sampling from $p^\star$ is intractable. We therefore learn a new parameterization $p_\theta(y)$ via amortized fine-tuning. Let $\mathcal{B}$ denote a replay buffer of sequences approximately sampled from $p^\star$ via tree search. The WDCE objective is:

$$\mathcal{L}_{\mathrm{WDCE}}(\theta) = \mathbb{E}_{y \sim \mathcal{B}}\, \mathbb{E}_{t, y_t \sim q_t(\cdot \mid y)} \Big[w(y) \sum_{\ell : (y_t)_\ell = [\mathrm{MASK}]} -\log p_\theta(y_\ell \mid (y_t)_{\mathrm{UM}}, t) \Big], \tag{26}$$

where the importance weight $w(y)$ is computed as:

$$w(y) = \frac{\exp(R(y; d^\star, x)/\alpha)}{\sum_{y' \in \mathcal{B}} \exp(R(y'; d^\star, x)/\alpha)}. \tag{27}$$

This softmax normalization over the buffer converts unnormalized log-weights into valid importance weights. Samples with $\kappa(y) = 0$ (non-binders) contribute no gradient.

To prevent collapse and preserve the pre-trained prior, we include a regularization term:

$$\mathcal{L}_{\mathrm{reg}}(\theta) = \mathrm{KL}(p_\theta \,\|\, p_{\theta_0}). \tag{28}$$

The full fine-tuning objective is:

$$\min_\theta \; \big\{ \mathcal{L}_{\mathrm{WDCE}}(\theta) + \lambda_{\mathrm{ctr}}\, \mathcal{L}_{\mathrm{ctr}}(\theta) + \lambda_{\mathrm{reg}}\, \mathcal{L}_{\mathrm{reg}}(\theta) \big\}, \tag{29}$$

where $\lambda_{\mathrm{ctr}}, \lambda_{\mathrm{reg}} \geq 0$ are hyperparameters. At generation time, users specify a target protein x and a desired direction $d^\star \in \{+1, -1\}$ corresponding to agonist or antagonist behavior.

3.6 DESIGN TASK

The directional allosteric design task is formally defined as:

Design Task: Given a desired direction $d^\star$, generate binder sequences whose induced transition asymmetry biases function toward the specified direction, without regressing kinetic rates or stabilizing endpoint states.

This formulation treats directionality of state transitions as the primary generative objective and defines a fully amortized procedure for incorporating coarse functional supervision into discrete sequence generation.

4 RESULTS

We designed experiments to test whether modeling binder action as a sequence-conditioned transition operator captures forms of allosteric control that are inaccessible to equilibrium- and structure-centric design methods. Our evaluation addresses three questions: (1) whether directionally fine-tuned generators induce non-reversible transition behavior; (2) whether directional control can be achieved independently of binding affinity; and (3) whether the framework supports targeted control over specific transition directions rather than global perturbations. Further case studies are shown in Appendix C.

4.1 EXPERIMENTAL SETTINGS

Dataset. We curated data from the IUPHAR/BPS Guide to Pharmacology database (Harding et al., 2026), classifying entries as agonist or antagonist based on bioactivity. Interacting residues on ligands were extracted using PeptiDerive (Sedan et al., 2016) and converted to canonical SMILES representations for downstream processing. To balance agonist and antagonist samples, we generated synthetic antagonist ligands via RFDiffusion (Watson et al., 2023) for targets with known agonists. After redundancy removal using MMseqs2 (Steinegger & Söding, 2017), the final dataset comprises 1,446 training and 336 held-out test samples.

The data split for TD3B training, validation, and test sets follows that of the Direction Oracle. Within the Direction Oracle training set, we further partitioned the data into training and validation subsets at an 8:1 ratio based on clustering. We filtered out binder sequences with residue counts outside the range of [16, 128], as well as targets containing only a single direction type (agonist or antagonist), to prevent direction bias toward specific targets. This preprocessing yields 130, 34, and 88 bidirectional target-binder pairs for the training, validation, and test sets, respectively.

Evaluation Metrics. For the Direction Oracle classification performance, we report Accuracy, Precision, Recall, and F1 Score to evaluate discriminative capability. To assess directional accuracy, inter-direction transitions, and designed binder affinity, we introduce direction-specific metrics for both agonist and antagonist modes: Affinity ($d^* = 1$), Affinity ($d^* = -1$), Direction Accuracy ($d^* = 1$), and Direction Accuracy ($d^* = -1$). Additionally, we report the gated reward to evaluate overall model training effectiveness.

Setup. We adopt the pre-trained MDLM weights from PepTune (Tang et al., 2025a) as our base model. During finetuning, we employ tree search for buffer generation following TR2-D2 (Tang et al., 2025c), regenerating binder data for multiple targets every k iterations with a first-in-first-out buffer to mitigate catastrophic forgetting. Binding affinity is estimated via a pre-trained predictor (Tang et al., 2025a; Zhang et al., 2026), trained on the PepLand dataset (Zhang et al., 2025) to produce a continuous, normalized affinity score (combining K_d, K_i, and IC_{50}), where higher values indicate stronger binding and a one-unit increase corresponds to an approximate tenfold change in binding strength. For the Direction Oracle, protein target sequences are encoded using ESM2 (Lin et al., 2023), and binder sequences are tokenized using the SPE tokenizer (Tang et al., 2025c). During generation, we apply Algorithm 2 for weighted sampling, generating 8 samples per specified direction for each target.

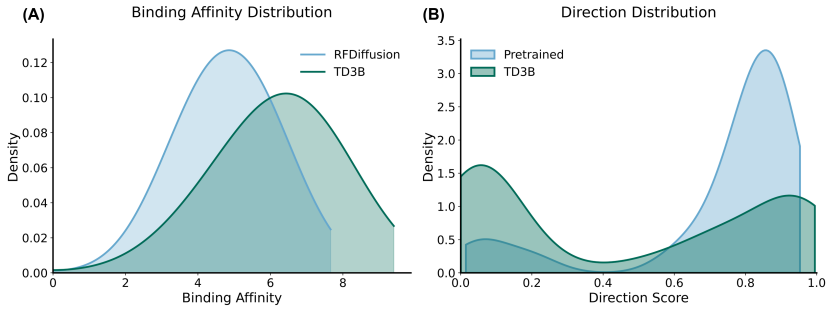

Figure 2: Direction and affinity distribution of TD3B. **(A)** Binding affinity comparison vs. RFDiffusion. **(B)** Direction comparison vs. pre-trained PepTune.

4.2 Direction Oracle for Protein-Peptide Binding

To ensure sufficient exploration space for both tree search and model training, we first demonstrate that the Direction Oracle can accurately identify the directionality of binders across diverse targets and varying sequence lengths, thereby providing reliable guidance during training. As shown in Table 1, the Direction Oracle achieves strong classification performance across all metrics.

Table 1: Binary classification performance of the Direction Oracle.

	Accuracy	Precision	Recall	F1
Dir. Oracle	0.93	0.90	0.91	0.90

4.3 Assessing Directional Asymmetry of Learned Transition Operators

Binding affinity is a prerequisite for characterizing agonists and antagonists (Brunton et al., 2018) To validate TD3B, we compared the predicted binding affinity of TD3B-generated samples against those designed by structure-based RFDiffusion (Watson et al., 2023). Figure 2A shows TD3B achieves higher predicted normalized affinity, demonstrating the effectiveness of gated reward fine-tuning.

We evaluated the directional behavior of peptides generated under agonist- versus antagonist-directed finetuning by measuring samples from the pre-trained (unconditioned) generator and TD3B using the Direction Oracle. As shown in Figure 2B, the pre-trained model generates predominantly agonist-biased binders with low confidence and lacks directional control. In contrast, TD3B produces distributions with higher confidence across both directions and enables explicit control over transition directionality, though we note asymmetry remains (see Table 2).

4.4 Comparison to Static and Predictive Baselines

We investigated whether TD3B can overcome the directional symmetry between activation and inhibition states. Since pre-trained discrete diffusion models do not support direction as a conditional input, we compare training-free guided diffusion approaches against tree-search-based finetuning. As shown in Table 2, TD3B achieves state-of-the-art performance on the gated reward, demonstrating its effectiveness for direction-specific generation. The directional accuracy shows asymmetry between agonist and antagonist modes, likely reflecting the stronger antagonist signal in the training data and the pre-trained model's inherent agonist bias. Finetuning-based methods achieve higher binding affinity while maintaining better balance across directions compared to guidance-only approaches. Compared to TR2-D2, TD3B employs weighted sampling to exploit high-potential samples and contrastive loss to separate sample distributions in latent space, enabling directional understanding beyond iterative optimization alone.

Table 2: Comparison of model performance across various metrics. Best results are in bold. CG: Classifier Guidance; SMC: Sequential Monte Carlo; TDS: Twisted Diffusion Sampler (Wu et al., 2023).

Model	Affinity		Dir. Acc.		Gated
	$(d^* = 1)$	$(d^* = -1)$	$(d^* = 1)$	$(d^* = -1)$	Reward
Pre-trained	4.80	4.80	0.84	0.18	2.1
CG	4.80	4.80	0.82	0.20	2.2
SMC	4.83	4.83	0.89	0.64	3.2
TDS	4.86	4.74	0.90	0.16	2.9
PepTune	4.94	4.94	0.88	0.28	2.8
TR2-D2	5.87	5.95	0.32	0.82	3.6
TD3B w/o $\mathcal{L}_{\text{ctr}}$	5.88	6.05	0.57	1.00	4.1
TD3B w/o $\mathcal{L}_{\text{reg}}$	5.82	5.46	**0.95**	0.30	3.8
TD3B	**5.92**	**6.30**	0.66	**1.00**	**5.2**

4.5 Targeted Control of Transition Asymmetry

We next assessed TD3B's targeted control over specific transitions without global dynamic perturbation. Conditioning on $d^\star \in \{+1, -1\}$, we verified that generated binders selectively bias transitions while maintaining high affinity. Table 3 shows de novo TD3B binders outperform length-matched wild-type references in affinity across all directions. Success rate—the fraction of binders achieving both (i) superior predicted affinity to wild-type and (ii) correct Direction Oracle classification—reaches 61% and 100% for forward and backward transitions, respectively. This confirms directionality as a controllable objective rather than an incidental outcome of binding optimization.

Table 3: Targeted control: Evaluation of functional specificity across transition objectives.

Design Objective (d^*)	**Affinity (WT)**	**Affinity (Transition)**	**Success Rate**
Forward Transition $(d^* = 1)$	4.66	5.81	0.61
Reverse Transition $(d^* = -1)$	4.99	6.06	1.00

4.6 Ablation Studies

Finally, we validate the effectiveness of each loss component. As shown in Table 2, removing either $\mathcal{L}_{ctr}$ or $\mathcal{L}_{reg}$ results in performance degradation. Without $\mathcal{L}_{ctr}$, direction accuracy drops substantially, as the overlapping latent space potentially undermines the model's ability to discriminate between activation and inhibition pathways. Moreover, without $\mathcal{L}_{reg}$, agonist accuracy increases to 0.95, yet generates biologically implausible sequences with impaired binding affinity, while antagonist mode nearly collapses (Acc 0.30), indicating that the antagonist feature space is more unstable to degradation during unconstrained finetuning.

5 Discussion

We introduce a generative framework, **TD3B**, that treats allosteric binder design as control over sequence-conditioned transition operators rather than optimization of static states or equilibrium energies. By making directional asymmetry and non-reversibility explicit modeling targets, the formulation captures functional effects that are not well-modeled by structure-centric design algorithms or predictive dynamics models. While we instantiate the framework with sequence-based binders, the transition-operator perspective is modality-agnostic and provides a general foundation for generative design in settings where function arises from non-equilibrium state shifts.

REFERENCES

Rahmad Akbar and Volkhard Helms. Allo: A tool to discriminate and prioritize allosteric pockets, 2018.

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

Marco Ballarotto, Sabine Willems, Tanja Stiller, Felix Nawa, Julian A Marschner, Francesca Grisoni, and Daniel Merk. De novo design of nurr1 agonists via fragment-augmented generative deep learning in low-data regime. *Journal of Medicinal Chemistry*, 66(12):8170–8177, 2023.

Gregory R Bowman, Eric R Bolin, Kathryn M Hart, Brendan C Maguire, and Susan Marqusee. Discovery of multiple hidden allosteric sites by combining markov state models and experiments. *Proceedings of the National Academy of Sciences*, 112(9):2734–2739, 2015.

Laurence L Brunton, Björn C Knollmann, Randa Hilal-Dandan, et al. *Goodman & Gilman's the pharmacological basis of therapeutics*, volume 13. McGraw-Hill Education New York, 2018.

Anne-Marinette Cao, Robert B Quast, Fataneh Fatemi, Philippe Rondard, Jean-Philippe Pin, and Emmanuel Margeat. Allosteric modulators enhance agonist efficacy by increasing the residence time of a gpcr in the active state. *Nature Communications*, 12(1):5426, 2021.

Hanqun Cao, Haosen Shi, Chenyu Wang, Sinno Jialin Pan, and Pheng-Ann Heng. GLID^{2}e: A gradient-free lightweight fine-tune approach for discrete biological sequence design. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=AHjspi4R22.

Tong Chen, Yinuo Zhang, and Pranam Chatterjee. Areuredi: Annealed rectified updates for refining discrete flows with multi-objective guidance. *arXiv preprint arXiv:2510.00352*, 2025a.

Tong Chen, Yinuo Zhang, Sophia Tang, and Pranam Chatterjee. Multi-objective-guided discrete flow matching for controllable biological sequence design. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*, 2025b. URL https://openreview.net/forum?id=8YIMLoHP9J.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Lewis A Churchfield, Annette Medina-Morales, Jeffrey D Brodin, Alfredo Perez, and F Akif Tezcan. De novo design of an allosteric metalloprotein assembly with strained disulfide bonds. *Journal of the American Chemical Society*, 138(40):13163–13166, 2016.

Declan Clarke, Anurag Sethi, Shantao Li, Sushant Kumar, Richard WF Chang, Jieming Chen, and Mark Gerstein. Identifying allosteric hotspots with dynamics: application to inter-and intra-species conservation. *Structure*, 24(5):826–837, 2016.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J Ragotte, Lukas F Milles, Basile IM Wicky, Alexis Courbet, Rob J de Haas, Neville Bethel, et al. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022.

Nina Dedic, Heather Dworak, Courtney Zeni, Grazia Rutigliano, and Oliver D Howes. Therapeutic potential of taar1 agonists in schizophrenia: evidence from preclinical models and clinical studies. *International Journal of Molecular Sciences*, 22(24):13185, 2021.

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

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *The Twelfth International Conference on Learning Representations*, 2024.

Najeeb Halabi, Olivier Rivoire, Stanislas Leibler, and Rama Ranganathan. Protein sectors: evolutionary units of three-dimensional structure. *Cell*, 138(4):774–786, 2009.

Simon D Harding, Jane F Armstrong, Elena Faccenda, Christopher Southan, Stephen PH Alexander, Anthony P Davenport, Michael Spedding, and Jamie A Davies. The iuphar/bps guide to pharmacology in 2026. *Nucleic Acids Research*, 54(D1):D1446–D1456, 2026.

Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450 (7172):964–972, 2007.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

Sijie Huang, Heng Liu, Yujin Wu, Joao M Braz, Divya Kranthi, Brendan W Hall, Xinyue Zhang, Dmytro S Radchenko, Yurii S Moroz, John J Irwin, et al. Structure based discovery of antipsychotic-like taar1 agonists. *bioRxiv*, pp. 2025–10, 2025.

Wenkang Huang, Shaoyong Lu, Zhimin Huang, Xinyi Liu, Linkai Mou, Yu Luo, Yanlong Zhao, Yaqin Liu, Zhongjie Chen, Tingjun Hou, et al. Allosite: a method for predicting allosteric sites. *Bioinformatics*, 29(18):2357–2359, 2013.

Eric M Jones, Nathan B Lubock, AJ Venkatakrishnan, Jeffrey Wang, Alex M Tseng, Joseph M Paggi, Naomi R Latorraca, Daniel Cancilla, Megan Satyadi, Jessica E Davis, et al. Structural and functional characterization of g protein–coupled receptors with deep mutational scanning. *Elife*, 9: e54895, 2020.

Lasse Karhu, Aniket Magarkar, Alex Bunker, and Henri Xhaard. Determinants of orexin receptor binding and activation—a molecular dynamics study. *The Journal of Physical Chemistry B*, 123, 2019.

Daniel E Koshland Jr, George Némethy, and David Filmer. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385, 1966.

Saifei Lei, Lachlan Clydesdale, Antao Dai, Xiaoqing Cai, Yang Feng, Dehua Yang, Yi-Lynn Liang, Cassandra Koole, Peishen Zhao, Thomas Coudrat, et al. Two distinct domains of the glucagon-like peptide-1 receptor control peptide-mediated biased agonism. *Journal of Biological Chemistry*, 293 (24):9370–9387, 2018.

Qiuzhen Li, Thomas Helleday, and Patrick Bryant. Rarefoldgpcr: Agonist design beyond natural amino acids. *bioRxiv*, pp. 2025–10, 2025.

Hui-Jun Liao and Jason T. C. Tzen. Investigating potential glp-1 receptor agonists in cyclopeptides from pseudostellaria heterophylla, linum usitatissimum, and drymaria diandra, and peptides derived from heterophyllin b for the treatment of type 2 diabetes: An in silico study. *Metabolites*, 2022.

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

Aaron Lou, Chenlin Meng, and Stefano Ermon. Discrete diffusion modeling by estimating the ratios of the data distribution. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 32819–32848, 2024.

Shaoyong Lu, Xinheng He, Duan Ni, and Jian Zhang. Allosteric modulator discovery: from serendipity to structure-based design. *Journal of medicinal chemistry*, 62(14):6405–6421, 2019.

Areesha Moiz, Kristian B Filion, Michael A Tsoukas, Oriana HY Yu, Tricia M Peters, and Mark J Eisenberg. The expanding role of glp-1 receptor agonists: a narrative review of current evidence and future directions. *EClinicalMedicine*, 86, 2025.

Jacque Monod, Jeffries Wyman, and Jean-Pierre Changeux. On the nature of allosteric transitions: a plausible model. *Journal of molecular biology*, 12(1):88–118, 1965.

Hesam N Motlagh, James O Wrabl, Jing Li, and Vincent J Hilser. The ensemble nature of allostery. *Nature*, 508(7496):331–339, 2014.

Seiji Nishino, Bonnie Ripley, Sebastiaan Overeem, Gert Jan Lammers, and Emmanuel Mignot. Hypocretin (orexin) deficiency in human narcolepsy. *The Lancet*, 355, 2000.

Hunter Nisonoff, Junhao Xiong, Stephan Allenspach, and Jennifer Listgarten. Unlocking guidance for discrete state-space diffusion and flow models. *arXiv preprint arXiv:2406.01572*, 2024.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, et al. One-shot design of functional protein binders with bindcraft. *Nature*, pp. 1–10, 2025.

Angus Phillips, Hai-Dang Dau, Michael John Hutchinson, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet. Particle denoising diffusion sampler. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 40688–40724. PMLR, 21–27 Jul 2024. URL `https://proceedings.mlr.press/v235/phillips24a.html`.

Arvind Pillai, Abbas Idris, Annika Philomin, Connor Weidle, Rebecca Skotheim, Philip JY Leung, Adam Broerman, Cullen Demakis, Andrew J Borst, Florian Praetorius, et al. De novo design of allosterically switchable protein assemblies. *Nature*, 632(8026):911–920, 2024.

Fabio Pirro, Nathan Schmidt, James Lincoff, Zachary X Widel, Nicholas F Polizzi, Lijun Liu, Michael J Therien, Michael Grabe, Marco Chino, Angela Lombardi, et al. Allosteric cooperation in a de novo-designed two-domain protein. *Proceedings of the National Academy of Sciences*, 117 (52):33246–33253, 2020.

Subham Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. *Advances in Neural Information Processing Systems*, 37:130136–130184, 2024.

Thomas E Scammell and Christopher J Winrow. Orexin receptors: pharmacology and therapeutic opportunities. *Annual Review of Pharmacology and Toxicology*, 51:243–266, 2011.

Yuval Sedan, Orly Marcu, Sergey Lyskov, and Ora Schueler-Furman. Peptiderive server: derive peptide inhibitors from protein–protein interactions. *Nucleic acids research*, 44(W1):W536–W541, 2016.

Qiancheng Shen, Guanqiao Wang, Shuai Li, Xinyi Liu, Shaoyong Lu, Zhongjie Chen, Kun Song, Junhao Yan, Lv Geng, Zhimin Huang, et al. Asd v3. 0: unraveling allosteric regulation with structural mechanisms and biological networks. *Nucleic acids research*, 44(D1):D527–D535, 2016.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis Titsias. Simplified and generalized masked diffusion for discrete data. *Advances in neural information processing systems*, 37: 103131–103167, 2024.

Diwakar Shukla, Yilin Meng, Benoît Roux, and Vijay S Pande. Activation pathway of src kinase reveals intermediate states as targets for drug design. *Nature communications*, 5(1):3397, 2014.

Hannes Stark, Felix Faltings, MinGyu Choi, Yuxin Xie, Eunsu Hur, Timothy John O'Donnell, Anton Bushuiev, Talip Uçar, Saro Passaro, Weian Mao, et al. Boltzgen: Toward universal binder design. *bioRxiv*, pp. 2025–11, 2025.

Martin Steinegger and Johannes Söding. Mmseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature biotechnology*, 35(11):1026–1028, 2017.

Zhen Wah Tan, Wei-Ven Tee, Enrico Guarnera, Lauren Booth, and Igor N Berezovsky. Allomaps: allosteric mutation analysis and polymorphism of signaling database. *Nucleic acids research*, 47 (D1):D265–D270, 2019.

Sophia Tang, Yinuo Zhang, and Pranam Chatterjee. Peptune: De novo generation of therapeutic peptides with multi-objective-guided discrete diffusion. In *Forty-second International Conference on Machine Learning*, 2025a. URL `https://openreview.net/forum?id=FQoy1Y1Hd8`.

Sophia Tang, Yinuo Zhang, Alexander Tong, and Pranam Chatterjee. Gumbel-softmax score and flow matching for discrete biological sequence generation. In *ICLR 2025 Workshop on AI for Nucleic Acids*, 2025b. URL https://openreview.net/forum?id=ITpCmDhSfu.

Sophia Tang, Yuchen Zhu, Molei Tao, and Pranam Chatterjee. Tr2-d2: Tree search guided trajectory-aware fine-tuning for discrete diffusion. *arXiv preprint arXiv:2509.25171*, 2025c.

Sophia Vincoff, Oscar Davis, Ismail Ilkan Ceylan, Alexander Tong, Joey Bose, and Pranam Chatterjee. SOAPIA: Siamese-guided generation of off target-avoiding protein interactions with high target affinity. In *ICML 2025 Workshop on Scaling Up Intervention Models*, 2025. URL https://openreview.net/forum?id=j0OpIG7leX.

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

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Thomas R Weikl and Fabian Paul. Conformational selection in protein binding and function. *Protein Science*, 23(11):1508–1518, 2014.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023.

Zheng Xu, Lulu Guo, Jingjing Yu, Siyuan Shen, Chao Wu, Weifeng Zhang, Chang Zhao, Yue Deng, Xiaowen Tian, Yuying Feng, et al. Ligand recognition and g-protein coupling of trace amine receptor taar1. *Nature*, 624(7992):672–681, 2023.

Jie Yin, Kerim Babaoglu, Chad A Brautigam, Lindsay Clark, Zhenhua Shao, Thomas H Scheuermann, Charles M Harrell, Anthony L Gotter, Anthony J Roecker, Christopher J Winrow, John J Renger, Paul J Coleman, and Daniel M Rosenbaum. Structure and ligand-binding mechanism of the human ox1 and ox2 orexin receptors. *Nature Structural & Molecular Biology*, 23(4):293–299, 2016.

Ruochi Zhang, Haoran Wu, Chang Liu, Qian Yang, Yuting Xiu, Kewei Li, Ningning Chen, Yu Wang, Yan Wang, Xin Gao, et al. Pepland: a large-scale pre-trained peptide representation model for a comprehensive landscape of both canonical and non-canonical amino acids. *Briefings in bioinformatics*, 26(4):bbaf367, 2025.

Yinuo Zhang, Sophia Tang, Tong Chen, Elizabeth Mahood, Sophia Vincoff, and Pranam Chatterjee. Peptiverse: A unified platform for therapeutic peptide property prediction. *bioRxiv*, pp. 2025–12, 2026.

Appendix

A ALGORITHMS

Algorithm 1 Direction-Only Amortized Fine-Tuning of an MDLM

Input: pre-trained MDLM $p_{\theta_0}(y)$
  directional dataset $\mathcal{D} = \{(y^{(n)}, a^{(n)})\}_{n=1}^{N}$
  direction oracle $f_\phi(y)$
  replay buffer $\mathcal{B} \leftarrow \emptyset$
Hyperparameters: learning rate η, temperature α, contrastive weight λ_{ctr}, KL weight λ_{reg}
while not converged **do**
  Sample minibatch $\{(y, a)\}$ from $\mathcal{D}$
  Compute direction labels $d(y) \in \{+1, -1\}$ and weights $\kappa(y)$
  Update direction oracle parameters ϕ using $\mathcal{L}_{\text{dir}}$
  ▷ *Populate replay buffer with direction-aligned samples*
  Sample candidate sequences $\{\tilde{y}_k\}_{k=1}^{M}$ from $p_\theta(y)$
  for each $\tilde{y}_k$ **do**
    Compute direction score $S(\tilde{y}_k) = \sigma(d^\star \cdot f_\phi(\tilde{y}_k)/\tau)$
    Set importance weight $w(\tilde{y}_k) \propto \exp(S(\tilde{y}_k)/\alpha)$
  end for
  Add weighted samples $\{(\tilde{y}_k, w(\tilde{y}_k))\}$ to replay buffer $\mathcal{B}$
  ▷ *Fine-tune MDLM using WDCE and contrastive objectives*
  Compute $\mathcal{L}_{\text{WDCE}}(\theta)$ using samples from $\mathcal{B}$
  Compute $\mathcal{L}_{\text{ctr}}(\theta)$ on non-negative samples
  Compute regularization $\mathcal{L}_{\text{reg}}(\theta) = \text{KL}(p_\theta \| p_{\theta_0})$
  Update $\theta \leftarrow \theta - \eta \nabla_\theta \big(\mathcal{L}_{\text{WDCE}} + \lambda_{\text{ctr}} \mathcal{L}_{\text{ctr}} + \lambda_{\text{reg}} \mathcal{L}_{\text{reg}}\big)$
end while
return fine-tuned generator $p_{\theta^*}(y)$ and oracle f_{ϕ^*}

Algorithm 2 Sampling Directional Allosteric Binders

Input: fine-tuned MDLM $p_{\theta^*}(y)$
  desired direction $d^\star \in \{+1, -1\}$
  direction oracle $f_{\phi^*}(y)$
  number of candidates M
Sample candidate binders $\{y_m\}_{m=1}^{M}$ i.i.d. from $p_{\theta^*}(y)$
for $m = 1$ **to** M **do**
  Compute direction score $S(y_m) = d^\star \cdot f_{\phi^*}(y_m)$
  Set weight $w_m \propto \exp(S(y_m))$
end for
Resample $y \sim \text{Categorical}(\{w_m\}_{m=1}^{M})$
return y ▷ *directionally biased binder*

B Theoretical Proofs

This appendix records basic guarantees for TD3B. We (i) justify exponential tilting as the unique solution of a KL-regularized improvement objective, (ii) characterize the population-optimal Direction Oracle under weighted logistic risk, (iii) relate weighted denoising cross-entropy to fitting a target (tilted) distribution, (iv) bound the effect of oracle approximation error on the induced tilted distribution, and (v) state a simple separability consequence of zero contrastive loss.

B.1 Notation and Setup

We adopt notation from the main text.

- $\mathcal{Y} := \mathcal{A}^L$ denotes the finite sequence space.
- $p_0(y)$ denotes a fixed base distribution on $\mathcal{Y}$ (e.g., the pre-trained MDLM distribution p_{θ_0}).
- $S : \mathcal{Y} \to \mathbb{R}$ denotes a score (objective) and $\alpha > 0$ a temperature.
- The (reward-)tilted distribution is

$$p^\star(y) := \frac{p_0(y)\exp(S(y)/\alpha)}{Z}, \qquad Z := \sum_{y' \in \mathcal{Y}} p_0(y')\exp(S(y')/\alpha)\,. \tag{30}$$

- For the direction task, we use labels $d(y) \in \{+1, -1\}$, confidence weights $\kappa(y) \in [0, 1]$, and an oracle $f_\phi : \mathcal{Y} \to \mathbb{R}$. The direction-only score used for design is $S(y; d^\star) = d^\star f_\phi(y)$ for $d^\star \in \{+1, -1\}$.
- For distributions P, Q on $\mathcal{Y}$ we write $\|P - Q\|_{\mathrm{TV}}$ for total variation and $\mathrm{KL}(P\|Q)$ for Kullback–Leibler divergence.

B.2 Exponential Tilting as KL-Regularized Improvement

Theorem B.1 (Exponential tilting solves KL-regularized improvement)**.** *Fix a base distribution p_0 on $\mathcal{Y}$ and a score $S : \mathcal{Y} \to \mathbb{R}$. Consider the optimization problem over distributions q on $\mathcal{Y}$:*

$$\max_{q \in \Delta(\mathcal{Y})} \left\{\mathbb{E}_{Y\sim q}[S(Y)] - \alpha\,\mathrm{KL}(q\|p_0)\right\}, \qquad \alpha > 0. \tag{31}$$

Then the unique maximizer is the tilted distribution $p^\star$ in equation 30.

Proof. Since $\mathcal{Y}$ is finite, equation 31 is a strictly concave optimization problem in q. Introduce a Lagrange multiplier λ for the constraint $\sum_y q(y) = 1$. The Lagrangian is

$$\mathcal{J}(q, \lambda) = \sum_y q(y)S(y) - \alpha \sum_y q(y)\log\frac{q(y)}{p_0(y)} + \lambda\left(\sum_y q(y) - 1\right). \tag{32}$$

Taking derivatives with respect to $q(y)$ and setting to zero gives

$$S(y) - \alpha\left(\log q(y) - \log p_0(y) + 1\right) + \lambda = 0, \tag{33}$$

so $\log q(y) = \log p_0(y) + S(y)/\alpha + c$ for a constant c. Hence

$$q(y) \propto p_0(y)\exp(S(y)/\alpha), \tag{34}$$

and normalization yields equation 30. Strict concavity implies uniqueness. □

Corollary B.1 (Limiting cases)**.** *Let $p^\star$ be defined by equation 30. Then:*

1. *As $\alpha \to \infty$, $p^\star \to p_0$ in total variation.*
2. *As $\alpha \to 0^+$, $p^\star$ concentrates on $\arg\max_{y\in\mathcal{Y}} S(y)$ (within the support of p_0).*

Proof. Write $p^\star(y) = p_0(y)\exp(S(y)/\alpha)/Z$. As $\alpha \to \infty$, $\exp(S(y)/\alpha) \to 1$ uniformly, so $Z \to 1$ and $p^\star \to p_0$. As $\alpha \to 0^+$, the normalization is dominated by maximizers of S because $\exp(S(y)/\alpha)$ is a softmax with temperature α. □

Proposition B.1 (Relative odds under tilting)**.** *For any $y_1, y_2 \in \mathcal{Y}$ with $p_0(y_1), p_0(y_2) > 0$,*

$$\frac{p^\star(y_1)}{p^\star(y_2)} = \frac{p_0(y_1)}{p_0(y_2)} \exp\left(\frac{S(y_1) - S(y_2)}{\alpha}\right). \tag{35}$$

In particular, if $p_0(y_1) = p_0(y_2)$ then the odds ratio depends only on $S(y_1) - S(y_2)$.

Proof. Immediate from the definition equation 30 since the normalizer Z cancels. □

Binary direction specialization. For $S(y; d^\star) = d^\star f(y)$ and $d^\star \in \{+1, -1\}$,

$$\frac{p^\star(y \mid +1)}{p^\star(y \mid -1)} = \frac{Z_-}{Z_+} \exp\left(\frac{2f(y)}{\alpha}\right), \qquad Z_\pm := \sum_y p_0(y)\exp(\pm f(y)/\alpha). \tag{36}$$

Thus larger oracle score $f(y)$ implies larger posterior odds of the $+1$-tilt relative to the -1-tilt.

B.3 Population Optimality of the Direction Oracle

We formalize the direction oracle as a weighted logistic risk minimizer.

Theorem B.2 (Bayes-optimal oracle under weighted logistic loss)**.** *Let (Y, D) be a random pair where $Y \in \mathcal{Y}$ and $D \in \{+1, -1\}$. Let $\kappa : \mathcal{Y} \times \{+1, -1\} \to [0, \infty)$ be a measurable weight (in TD3B, κ encodes down-weighting of partial agonists and exclusion of non-binders). Consider minimizing*

$$\mathcal{R}(f) := \mathbb{E}[\kappa(Y, D)\ \log(1 + \exp(-Df(Y)))] \tag{37}$$

over all functions $f : \mathcal{Y} \to \mathbb{R}$. Define

$$\eta_+(y) := \mathbb{E}[\kappa(Y, D)\mathbf{1}\{D = +1\} \mid Y = y], \qquad \eta_-(y) := \mathbb{E}[\kappa(Y, D)\mathbf{1}\{D = -1\} \mid Y = y]. \tag{38}$$

If $\eta_+(y) + \eta_-(y) > 0$, the pointwise minimizer satisfies

$$f^\star(y) = \log \frac{\eta_+(y)}{\eta_-(y)}. \tag{39}$$

Proof. Fix $y \in \mathcal{Y}$ and write the conditional risk (up to an additive constant independent of $f(y)$) as

$$r_y(u) = \eta_+(y)\log(1 + \exp(-u)) + \eta_-(y)\log(1 + \exp(u)), \qquad u := f(y). \tag{40}$$

This is strictly convex in u. Differentiate and set to zero:

$$r_y'(u) = -\eta_+(y)\frac{1}{1 + \exp(u)} + \eta_-(y)\frac{\exp(u)}{1 + \exp(u)} = 0. \tag{41}$$

Multiplying by $1 + \exp(u)$ gives $-\eta_+(y) + \eta_-(y)\exp(u) = 0$, hence $\exp(u) = \eta_+(y)/\eta_-(y)$ and equation 39 follows. □

Remark. Theorem B.2 shows that, in the population limit, a weighted logistic oracle estimates a (weighted) log-odds function. This makes exponential tilting with $S(y; d^\star) = d^\star f(y)$ a principled way to bias generation toward one directional class.

B.4 Weighted Denoising Cross-Entropy Fits a Target Distribution

We formalize the effect of WDCE as fitting an MDLM to a reweighted data distribution.

Lemma B.1 (Weighted risk equals unweighted risk under a reweighted distribution)**.** *Let r be any distribution on $\mathcal{Y}$ and let $w : \mathcal{Y} \to [0, \infty)$ be a weight function with $0 < \mathbb{E}_{Y\sim r}[w(Y)] < \infty$. Define the normalized reweighted distribution*

$$\pi(y) := \frac{r(y)w(y)}{\mathbb{E}_{Y\sim r}[w(Y)]}. \tag{42}$$

Then for any nonnegative loss $\ell(y)$,

$$\mathbb{E}_{Y\sim r}[w(Y)\ell(Y)] = \mathbb{E}_{Y\sim r}[w(Y)] \cdot \mathbb{E}_{Y\sim\pi}[\ell(Y)]. \tag{43}$$

Proof. By definition, $\mathbb{E}_{Y\sim\pi}[\ell(Y)] = \sum_y \pi(y)\ell(y) = \frac{1}{\mathbb{E}_r[w]} \sum_y r(y)w(y)\ell(y)$. Rearrange. □

Theorem B.3 (Population optimality of WDCE denoisers)**.** *Fix a corruption kernel family $q_t(x_t \mid x)$ and a time sampling distribution over $t \in [0, 1]$. Let π be a target distribution on $\mathcal{Y}$ and define the joint (X, X_t) by $X \sim \pi$ and $X_t \sim q_t(\cdot \mid X)$. Consider the MDLM denoising objective*

$$\mathcal{L}_\pi(\theta) = \mathbb{E}_{t,X,X_t}\left[\sum_{\ell:(X_t)_\ell=[\mathrm{MASK}]} -\log p_\theta(X_\ell \mid (X_t)_{\mathrm{UM}}, t)\right]. \tag{44}$$

Then for each time t and each masked position ℓ, the minimizer satisfies

$$p_{\theta^\star}(X_\ell = a \mid (X_t)_{\mathrm{UM}}, t) = \mathbb{P}_\pi(X_\ell = a \mid (X_t)_{\mathrm{UM}}, t) \qquad \forall a \in \mathcal{A}, \tag{45}$$

that is, the optimal denoiser recovers the true conditional under π.

Proof. Fix t and condition on the context $C := ((X_t)_{\mathrm{UM}}, t)$ and the event that position ℓ is masked. The inner term in equation 44 is the cross-entropy between the true conditional distribution of X_ℓ given C and the model distribution $p_\theta(\cdot \mid C)$. Cross-entropy is minimized uniquely by matching the true conditional. Taking expectation over contexts yields the result. □

Connection to reward tilting. If the proposal distribution r is p_0 and weights are $w(y) = \exp(S(y)/\alpha)$, then the reweighted distribution π in Lemma B.1 equals the tilted distribution $p^\star$ in equation 30. Thus WDCE is (in the population limit) standard MDLM training under the tilted target distribution.

B.5 Stability of Tilting Under Oracle Approximation Error

Theorem B.4 (Tilt robustness under bounded score error)**.** *Let $S^\star : \mathcal{Y} \to \mathbb{R}$ be an ideal score and let $S : \mathcal{Y} \to \mathbb{R}$ satisfy*

$$\sup_{y\in\mathcal{Y}} |S(y) - S^\star(y)| \le \varepsilon. \tag{46}$$

Let $p^\star$ and $\widetilde{p}$ be the corresponding tilted distributions built from $(p_0, S^\star)$ and (p_0, S) with the same temperature $\alpha > 0$. Then

$$\mathrm{KL}(\widetilde{p}\|p^\star) \le \frac{2\varepsilon}{\alpha}, \qquad \mathrm{KL}(p^\star\|\widetilde{p}) \le \frac{2\varepsilon}{\alpha}, \tag{47}$$

and therefore, by Pinsker's inequality,

$$\|\widetilde{p} - p^\star\|_{\mathrm{TV}} \le \sqrt{\frac{\varepsilon}{\alpha}}. \tag{48}$$

Proof. Write $S = S^\star + \delta$ with $|\delta(y)| \le \varepsilon$. Then

$$\widetilde{p}(y) = \frac{p_0(y)\exp((S^\star(y) + \delta(y))/\alpha)}{\widetilde{Z}} = p^\star(y)\,\frac{\exp(\delta(y)/\alpha)}{\mathbb{E}_{Y\sim p^\star}[\exp(\delta(Y)/\alpha)]}. \tag{49}$$

Since $\exp(\delta/\alpha) \in [\exp(-\varepsilon/\alpha), \exp(\varepsilon/\alpha)]$, the normalizer ratio satisfies

$$\mathbb{E}_{p^\star}[\exp(\delta/\alpha)] \in [\exp(-\varepsilon/\alpha), \exp(\varepsilon/\alpha)]. \tag{50}$$

Hence for all y,

$$\log\frac{\widetilde{p}(y)}{p^\star(y)} = \frac{\delta(y)}{\alpha} - \log \mathbb{E}_{p^\star}[\exp(\delta/\alpha)] \in \left[-\frac{2\varepsilon}{\alpha}, \frac{2\varepsilon}{\alpha}\right]. \tag{51}$$

Taking expectation under $\widetilde{p}$ yields $\mathrm{KL}(\widetilde{p}\|p^\star) \le 2\varepsilon/\alpha$. The reverse KL bound follows symmetrically by swapping roles of $(S, S^\star)$. Pinsker's inequality gives the TV bound. □

B.6 A Separability Consequence of Zero Contrastive Loss

Proposition B.2 (Zero margin-contrastive loss implies linear separability)**.** *Let $\{(y_i, d_i)\}_{i=1}^N$ be labeled samples with $d_i \in \{+1, -1\}$ and embeddings $h(y_i) \in \mathbb{R}^m$. Consider the margin-contrastive loss*

$$\mathcal{L}_{\mathrm{ctr}} = \sum_{(i,j):d_i=d_j} \|h(y_i) - h(y_j)\|_2^2 + \sum_{(i,j):d_i\neq d_j} \max(0,\, m_0 - \|h(y_i) - h(y_j)\|_2)^2. \tag{52}$$

If $\mathcal{L}_{\mathrm{ctr}} = 0$, then there exist $u_+, u_- \in \mathbb{R}^m$ such that $h(y_i) = u_{d_i}$ for all i and $\|u_+ - u_-\|_2 \ge m_0$. In particular, the classes are linearly separable by a hyperplane with margin at least $m_0/2$.

Proof. If $\mathcal{L}_{\mathrm{ctr}} = 0$, then for any pair (i, j) with $d_i = d_j$ we must have $\|h(y_i) - h(y_j)\|_2^2 = 0$, hence all embeddings within a class are identical. Denote the two class prototypes by u_+ and u_-. For any pair with $d_i \neq d_j$, the hinge term being zero implies $\|u_+ - u_-\|_2 \ge m_0$.

Define $w := u_+ - u_-$ and $b := -\frac{1}{2}\langle w, u_+ + u_-\rangle$. Then

$$\langle w, u_+\rangle + b = \tfrac{1}{2}\|u_+ - u_-\|_2^2 \ge \tfrac{1}{2}m_0^2, \qquad \langle w, u_-\rangle + b = -\tfrac{1}{2}\|u_+ - u_-\|_2^2 \le -\tfrac{1}{2}m_0^2, \tag{53}$$

so the hyperplane $\{z : \langle w, z\rangle + b = 0\}$ separates the two prototypes. The (geometric) margin is at least $\|u_+ - u_-\|_2/2 \ge m_0/2$. □

C CASE STUDIES

We fine-tuned TD3B on two key GPCR targets (GLP1R and TAAR1) and selected top-ranked candidates for each. Complex structures were predicted using AlphaFold3, followed by binding-site detection and scoring with an external classifier. We first focused on GLP-1R, a clinically important metabolic receptor revolutionary for obesity, weight loss, and type 2 diabetes treatment where agonist activity is the primary therapeutic mechanism (Moiz et al., 2025). As reference, Figure C1A shows the AF3-predicted structure of GLP-1R with all 37 interacting residues at the binding interface of the endogenous GLP-1 hormone. TD3B-designed agonists engage key activation residues including Tyr148, Tyr152, Arg190, Lys197, Tyr205, Gln234, Trp297, Thr298, Arg299, and Asn300 (Liao & Tzen, 2022), with Arg299 and Asn300 being essential for full agonist activity (Lei et al., 2018) (Figure C1B). In contrast, TD3B-designed antagonists (Figure C1C) lack interactions with Arg299 and Asn300, consistent with their inability to activate the receptor. These results demonstrate that TD3B can design agonist and antagonist binders that selectively engage or avoid critical activation residues on the same receptor.

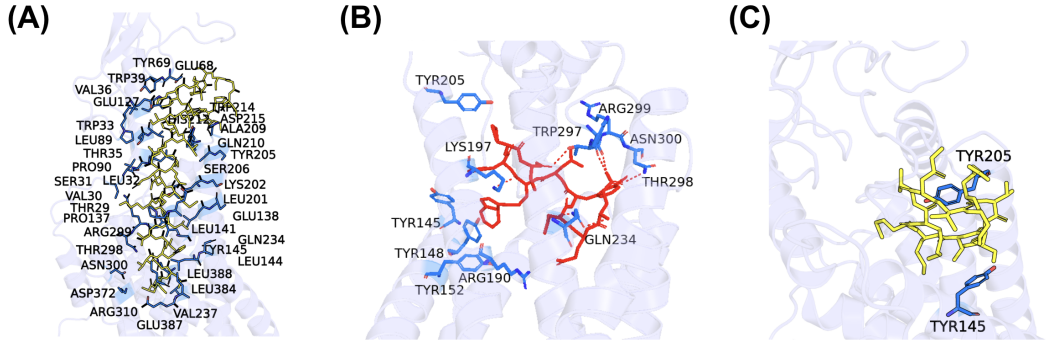

Figure C1: Evaluation of TD3B on GLP-1R. **(A)** Existing GLP-1 peptide hormone bound to GLP-1R. **(B)** TD3B-designed agonist bound GLP-1R. **(C)** TD3B-designed antagonist bound GLP-1R.

We next applied TD3B to the Orexin 1 Receptor (OX1R), a regulatory GPCR where antagonists treat insomnia and agonists show promise for narcolepsy (Scammell & Winrow, 2011; Nishino et al., 2000). Figure C2A shows the crystal structure of OX1R bound to the clinical antagonist suvorexant (PDB: 4ZJ8), revealing 14 key orthosteric binding site residues. TD3B-designed agonists engage 5 conserved residues, while antagonists engage 6 residues (Figures C2B and C). Critically, both ligands interact with GLN126, a molecular switch that controls receptor activation through hydrogen bonding with TYR348 (Karhu et al., 2019). In antagonists, the GLN126-TYR348 bond stabilizes the inactive state, while agonists are predicted to disrupt this interaction to enable activation (Yin et al., 2016). The substantial conservation with suvorexant's validated binding site confirms orthosteric targeting and demonstrates TD3B's ability to design functionally distinct ligands for the same receptor pocket. Together, these results indicate that TD3B can generate functionally divergent binders for the same target by modulating transition directionality rather than binding affinity alone. More cases are shown in Appendix C.

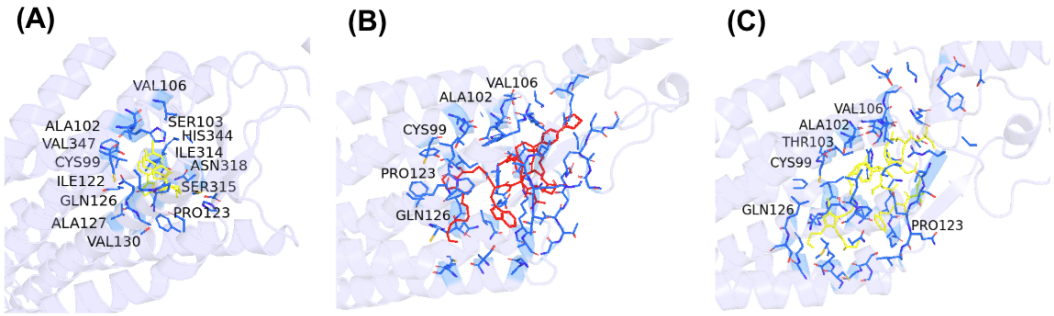

Figure C2: Evaluation of TD3B on OX1R. **(A)** Suvorexant bound to OX1R (PDB: 4ZJ8). **(B)** TD3B-designed agonist. **(C)** TD3B-designed antagonist.

We further applied TD3B to TAAR1, a neuromodulatory GPCR implicated in dopaminergic and serotonergic signaling and strongly linked to schizophrenia, where both agonists and antagonists are of pharmacological interest (Dedic et al., 2021). As shown in Figure C3, TD3B-generated agonists and antagonists for TAAR1 exhibit distinct binding modes. Similar to the activatory contacts of T1AM (3-iodothyronamine), a small molecule TAAR1 agonist shown in PDB 8JLN (Figure C3A) (Xu et al., 2023), the designed agonist preferentially engages with a compact orthosteric pocket spanning TM3, TM5, TM6, and TM7, (Jones et al., 2020) including residues associated with receptor activation (Huang et al., 2025) (Figure C3B), and the designed antagonist occupies a broader binding region that partially overlaps the orthosteric site but selectively avoids key activation-associated transmembrane contacts (Figure C3C).

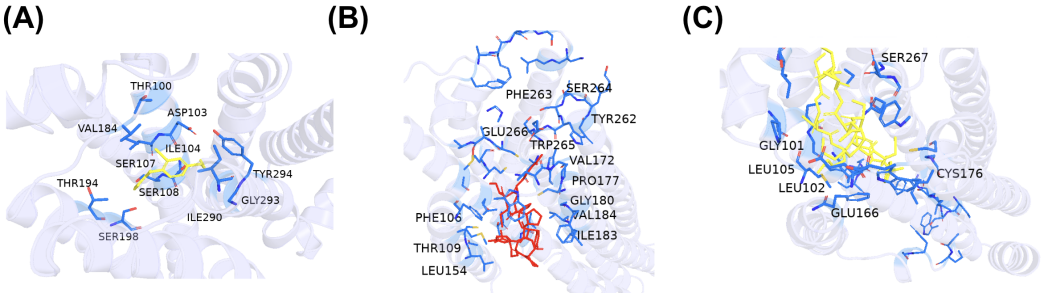

Figure C3: Evaluation of TD3B on TAAR1. **(A)** Existing TAAR1 agonist bound to TAAR1. **(B)** TD3B-designed agonist bound TAAR1. **(C)** TD3B-designed antagonist bound TAAR1.

D Related Works

D.1 Allosteric Binder Design

Classical allosteric theory originated from the MWC concerted transition (Monod et al., 1965) and KNF sequential induced-fit (Koshland Jr et al., 1966) models, while early discovery relied on serendipitous screening hits (Lu et al., 2019). With the expansion of structural databases (Shen et al., 2016), computational methods emerged to predict allosteric sites (Huang et al., 2013; Akbar & Helms, 2018), hotspots (Clarke et al., 2016), and communication pathways (Tan et al., 2019; Halabi et al., 2009) from static structures. However, static approaches miss cryptic sites that remain occluded in certain conformations. Dynamics-based methods using MD simulations and Markov state models (Shukla et al., 2014; Bowman et al., 2015) address this limitation but incur high computational costs.

De novo allosteric design has evolved from early Rosetta-based side-chain networks (Churchfield et al., 2016; Pirro et al., 2020) to recent modular rigid-body coupling strategies leveraging RFDiffusion (Watson et al., 2023) and ProteinMPNN (Dauparas et al., 2022), enabling peptide-responsive ring architectures and effector-induced cage disassembly (Pillai et al., 2024). Complementary approaches include structure-based design for GPCRs (Li et al., 2025) and Chemical Language Model (CLM)-based generation (Ballarotto et al., 2023). In contrast, TD3B reframes directional design as a non-equilibrium transport problem, introducing directional supervision such that binding agents act as one-way valves controlling state transition directionality.

D.2 Conditional Generation for Discrete Diffusion

Diffusion models have emerged as powerful unsupervised generative frameworks capable of capturing distributions over discrete spaces, with conditional generation enabling efficient sample exploration (Austin et al., 2021; Lou et al., 2024; Sahoo et al., 2024; Shi et al., 2024). To steer generation toward desired properties, guided generation approaches incorporate classifier gradients into the sampling process. These include training-free methods such as Classifier Guidance (CG) and Sequential Monte Carlo (SMC) (Dhariwal & Nichol, 2021; Nisonoff et al., 2024; Chung et al., 2022; Wu et al., 2023; Dou & Song, 2024; Phillips et al., 2024), as well as Classifier-Free Guidance (CFG) (Ho & Salimans, 2022). While computationally efficient, these approaches struggle to provide complex guidance in discrete domains, where gradient-based steering is inherently limited by non-differentiable operations.

Reinforcement learning-based methods address this limitation by enabling discrete diffusion models to learn more sophisticated conditional distributions through environment interaction. Policy gradient approaches such as DRAKES and GLID^{2}E (Wang et al., 2025; Cao et al., 2025) update model policies via reward signals, while tree-search methods including PepTune and TR2-D2 (Tang et al., 2025a;c) offer greater flexibility for multi-condition and multi-objective sampling by explicitly exploring the combinatorial space.

E IMPLEMENTATION DETAILS

E.1 DIRECTION ORACLE TRAINING

The Direction Oracle is trained for 20 epochs using AdamW optimization with a learning rate of 10^{-5} and batch size of 16, minimizing cross-entropy loss. pre-trained encoders remain frozen throughout training; only projection layers, self-attention and cross-attention modules, and the two-layer MLP classifier head are optimized. Due to the limited size of available labeled data, we train on the full training split without validation and evaluate performance exclusively on an independent held-out test set.

E.2 TD3B FINETUNING

For tree search-based sampling, we employ trajectory-aware tree search with 20 iterations and 24 children per node, sampling 4 targets per iteration with a buffer size of 32 candidates per target. A replay buffer of 2000 samples with FIFO replacement is maintained to mitigate catastrophic forgetting. Training uses a batch size of 4 with gradient accumulation over 4 steps, a learning rate of 5×10^{-5}, and 4 WDCE replicates per sample. The KL regularization coefficient λ_{reg} is set to 0.5, and tree search resampling is performed every 10 epochs. Training is conducted on 8 NVIDIA A100 GPUs using PyTorch DDP with synchronized buffer aggregation.