HYBRID DIFFUSION–SCHRÖDINGER DYNAMICS WITH UNIFIED LEARNING FOR TARGET-CONDITIONED PEPTIDE CO-DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Target-conditioned peptide co-design is inherently challenging, as the generated peptides must exhibit biologically plausible sequences and structurally valid conformations. This difficulty is further amplified by the strong constraints imposed by the target protein, requiring physicochemical compatibility and geometric complementarity with the binding interface. Under these coupled constraints, existing methods often struggle to generate bindable peptides, resulting in limited binding success rates. In this paper, we propose **PepHyD**, a hybrid generative dynamics framework that formulates peptide co-design as a transition from broad sequence–structure exploration to target-directed transport. Specifically, PepHyD adapts the generative dynamics to the evolving state of generation, transitioning from diffusion-driven exploration of diverse peptide configurations to Schrödinger bridge transport that progressively concentrates intermediate representations toward target-compatible states. This design accommodates early-stage uncertainty while progressively enforcing target-specific constraints. Moreover, we introduce a unified target parameterization for both dynamics, allowing them to be learned within a shared denoising framework. Extensive experiments on LNR and BioDB demonstrate state-of-the-art performance, while evaluations on unseen targets and binding pockets further show strong generalization.

1 INTRODUCTION

Peptides have emerged as an attractive class of therapeutic and molecular recognition agents due to their high binding specificity, favorable biocompatibility Gare et al. (2025), and ability to engage protein surfaces that are often difficult to target with small molecules Xiao et al. (2025). These properties have supported their use in diverse therapeutic applications, including cancer, metabolic disorders, and immune-related diseases Yang et al. (2025); Stein Gold et al. (2025). Traditional peptide design typically relies on combinatorial library screening, molecular docking, and physics-based optimization Chen et al. (2024), which are often time-consuming and resource-intensive, limiting their scalability for peptide discovery Ekambaram & Dokholyan (2026). This motivates the application of advanced deep learning techniques to peptide design, offering an efficient approach to discovering promising peptides and accelerating the development of peptide therapeutics.

Existing peptide design approaches can be broadly categorized into sequence generation, structure generation, and joint sequence–structure design Wu et al. (2024); Wang et al. (2025a). Sequence-based methods primarily learn biochemical patterns among amino acids, without guaranteeing that the resulting sequences can adopt structurally valid conformations Chen et al. (2026). Structure-based methods focus on generating peptide backbones, with amino-acid sequences typically assigned in a separate design stage Watson et al. (2023); Dauparas et al. (2022). This separation may overlook the intrinsic interdependence between sequence and structure, limiting sequence–structure compatibility. Consequently, sequence–structure co-design has emerged as an important paradigm for jointly generating amino-acid identities and 3D conformations Kong et al. (2024); Li et al. (2024b), but remains subject to stringent constraints imposed by the binding target. First, the generated sequence must exhibit physicochemical compatibility with the target interface, including appropriate hydrophobic, electrostatic, and intermolecular interaction patterns. Second, the peptide conformation must closely match the geometry of the target binding site, since even small deviations

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

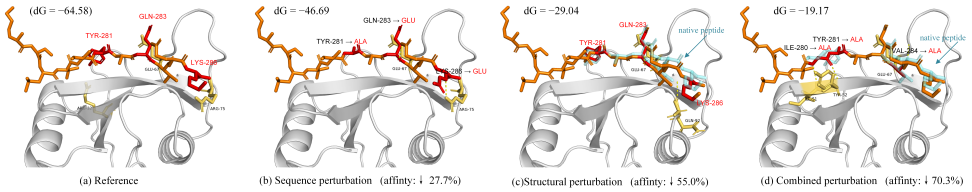


Figure 1: Binding sensitivity to local perturbations. (a) Native complex. (b) Sequence perturbation alters interface physicochemical properties; for example, replacing Lys^+ with Glu^- reduces the binding-energy magnitude by 27.7%. (c) Structural perturbation weakens binding by 55.0%. (d) Combined perturbation further increases the reduction to 70.3%. These results highlight the joint importance of physicochemical compatibility and geometric complementarity at the target interface.

in backbone or side-chain arrangement can disrupt key binding interactions. Importantly, these two requirements are strongly coupled: residue identities shape both interactions and conformational preferences, while the resulting structure determines their arrangement at the interface, as shown in Fig. 1. Therefore, designing target-compatible peptides remains challenging.

Recently, deep generative models Oussidi & Elhassouny (2018), particularly diffusion Ho et al. (2020) and flow-based methods Lipman et al. (2022), have been widely adopted for target-conditioned peptide generation Wang et al. (2025b); Du et al. (2024). Diffusion models progressively perturb data with noise and learn the reverse denoising process Liu et al. (2024), whereas Flow Matching learns a continuous transport path between distributions Liu et al. (2022). Those generative processes can be viewed as transforming a simple prior distribution, typically Gaussian, into increasingly structured peptide representations, progressing from coarse backbone configurations toward clean peptide data. A key challenge is how to transform these coarse representations into configurations compatible with the target binding interface. Fig. 1 shows that small local perturbations in residue properties or geometry can substantially disrupt peptide–target interactions, indicating successful binding depends on accurately resolving fine-grained physicochemical and geometric constraints. Therefore, a strong target-conditioned connection should be established between coarse sequence–structure organization and clean, target-compatible states, enabling peptide representations to progressively evolve under the constraints imposed by the binding interface.

To tackle these challenges, we propose **PepHyD**, a hybrid Diffusion–Schrödinger bridge dynamics framework with unified learning for target-conditioned peptide co-design. PepHyD organizes the generative trajectory according to the evolving requirements of peptide generation, progressing from broad sequence–structure exploration to increasingly target-constrained generation. Specifically, starting from highly uncertain latent states, the trajectory is initially governed by diffusion dynamics to explore diverse peptide configurations and establish coarse sequence–structure organization. As the peptide representations become more structured, the dynamics transition to Schrödinger bridge transport, directing the remaining trajectory toward target-compatible states under binding-interface constraints. The key reason is that Schrödinger bridges model stochastic transport between paired endpoints, coupling intermediate peptide states with target-compatible clean states. This allows the model to learn how sequence and geometry should evolve from coarse organization toward binding-compatible configurations. Moreover, we introduce a unified target parameterization that accommodates both stochastic dynamics under a common learning formulation, enabling them to share the same prediction target and training objective. In summary, our key contributions include:

- We highlight a fundamental challenge: even small local deviations in residue identity or geometry can substantially disrupt peptide–target interactions, underscoring the need for both physicochemical compatibility and geometric complementarity at the binding interface. This provides an important consideration for generative modeling in target-conditioned peptide co-design.
- We propose a novel hybrid generative dynamics framework, **PepHyD**, in which the stochastic dynamics adapt to evolving peptide representations. The process transitions from diffusion-driven exploration to Schrödinger bridge transport toward target-compatible configurations, mitigating the common issue that structurally plausible peptides may still exhibit poor target binding.
- PepHyD achieves state-of-the-art binding performance on LNR and BioDB, with success rates of 85.05% and 87.01%, respectively. It further attains an average success rate of 86.7% across multiple pockets of an unseen therapeutically relevant target, while molecular dynamics simulations demonstrate improved binding stability and interface preservation.

2 RELATED WORK

2.1 GENERATIVE MODELS FOR PEPTIDE CO-DESIGN

Recent advances in deep generative modeling have substantially improved target-conditioned peptide co-design by jointly modeling peptide sequences and three-dimensional structures. PepGLAD introduces geometric latent diffusion for variable-size full-atom peptide generation (Kong et al., 2024), while PepTri incorporates physics-based, evolutionary, and mutual-information guidance into latent diffusion to enhance sequence–structure generation (Nguyen et al., 2026). Beyond diffusion, PepFlow formulates full-atom peptide design as multimodal flow matching over residue types, backbone frames, and side-chain torsions (Li et al., 2024b). PepHAR adopts a hotspot-driven autoregressive strategy that progressively extends peptide fragments into complete binders (Li et al., 2024a). PepBFN employs Riemannian–Euclidean Bayesian flows to jointly model residue identities and structural variables (Qian et al., 2026). However, these methods often struggle to balance sequence–structure plausibility with target-specific binding requirements. Our work instead explores hybrid stochastic dynamics within a unified trajectory, using diffusion for broad exploration and Schrödinger bridge transport toward target-compatible peptide states.

2.2 SCHRÖDINGER BRIDGE GENERATIVE MODELING

Schrödinger bridges provide a stochastic optimal transport framework for connecting prescribed endpoint distributions while remaining close to a reference process. Recent studies have explored this formulation across a variety of generative tasks. Diffusion Schrödinger Bridge (DSB) connects Schrödinger bridge optimization with score-based generative modeling through an iterative procedure for learning the associated forward and backward dynamics (De Bortoli et al., 2021). I^2SB extends Schrödinger bridge modeling to image-to-image translation by constructing stochastic transport between paired source and target images, with applications including inpainting, super-resolution, deblurring, and JPEG restoration (Liu et al., 2023). For unpaired image translation, UNSB formulates the bridge through a sequence of adversarial learning problems and extends Schrödinger bridge modeling to high-resolution translation across image domains (Kim et al., 2024). These studies demonstrate the effectiveness of Schrödinger bridge modeling for conditional stochastic transport. Building on this idea, our work extends it to molecular design, using endpoint-conditioned bridge dynamics to progressively steer peptide representations toward regions compatible with the target binding interface.

3 METHODS

We represent the target binding site and peptide as residue-level graphs $\mathcal{G}^b$ and $\mathcal{G}^p$, respectively, and perform generation in a continuous latent space. Details of the graph construction and model architecture are provided in Appendix B.1. PepHyD adapts the stochastic dynamics along the generative trajectory: diffusion dynamics support exploration from highly uncertain latent states, while Schrödinger bridge (SB) dynamics progressively constrain the evolving representation toward target-compatible peptide states. To formalize these dynamics, we first introduce the stochastic formulations of score-based diffusion models and Schrödinger bridges (SBs). Let $\mathbf{X}_t \in \mathbb{R}^d$ denote a continuous-time stochastic process for $t \in [0, 1]$, with W_t and $\overline{W}_t$ denoting the forward- and reverse-time Wiener processes, respectively.

3.1 DIFFUSION AND SCHRÖDINGER BRIDGE DYNAMICS

Diffusion Dynamics. Score-based diffusion models define a stochastic process that gradually perturbs the data distribution toward a tractable prior and reverses this transformation through the corresponding reverse-time dynamics. The forward and backward SDEs are given by:

$$d\mathbf{X}_t = f_t(\mathbf{X}_t)\, dt + \sqrt{\beta_t}\, dW_t, \qquad \mathbf{X}_0 \sim p_{\text{data}}, \tag{1}$$

$$d\mathbf{X}_t = [f_t(\mathbf{X}_t) - \beta_t \nabla_{\mathbf{x}} \log p_t(\mathbf{X}_t)]\, dt + \sqrt{\beta_t}\, d\overline{W}_t. \tag{2}$$

where f_t is the drift function, $\beta_t > 0$ controls the diffusion strength, and $p_t(\mathbf{x})$ denotes the marginal density of $\mathbf{X}_t$. The forward dynamics are typically designed such that $\mathbf{X}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The term $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ denotes the score function, which is parameterized by a neural network and learned through denoising score matching Song et al. (2021).

Schrödinger Bridge Dynamics. Score-based diffusion models typically terminate at a predefined prior, limiting the flexibility of the terminal boundary. Schrödinger bridges (SBs) generalize this setting to stochastic transport between two arbitrary boundary distributions p_A and p_B. Given a reference path measure $\mathbb{P}$, the SB seeks a path measure $\mathbb{Q}^\star$ that minimizes the KL divergence from $\mathbb{P}$ while satisfying the prescribed endpoint marginals:

$$\mathbb{Q}^\star = \arg\min_{\mathbb{Q}\in\mathcal{P}(p_A,p_B)} D_{\mathrm{KL}}\left(\mathbb{Q}\,\|\,\mathbb{P}\right), \tag{3}$$

where $\mathcal{P}(p_A, p_B)$ denotes the set of path measures with initial and terminal marginals p_A and p_B, respectively. The optimal bridge is characterized by the following forward and backward SDEs:

$$d\mathbf{X}_t = \left[f_t(\mathbf{X}_t) + \beta_t \nabla_{\mathbf{x}} \log \Psi(\mathbf{X}_t, t)\right] dt + \sqrt{\beta_t}\, dW_t, \qquad \mathbf{X}_0 \sim p_A, \tag{4}$$

$$d\mathbf{X}_t = \left[f_t(\mathbf{X}_t) - \beta_t \nabla_{\mathbf{x}} \log \widehat{\Psi}(\mathbf{X}_t, t)\right] dt + \sqrt{\beta_t}\, d\overline{W}_t, \qquad \mathbf{X}_1 \sim p_B, \tag{5}$$

where Ψ and $\widehat{\Psi}$ are the time-dependent Schrödinger potentials. They satisfy the coupled equations

$$\begin{cases} \dfrac{\partial \Psi(\mathbf{x}, t)}{\partial t} = -\nabla \Psi(\mathbf{x}, t)^\top f_t(\mathbf{x}) - \dfrac{1}{2}\beta_t \Delta \Psi(\mathbf{x}, t), \\ \dfrac{\partial \widehat{\Psi}(\mathbf{x}, t)}{\partial t} = -\nabla \cdot \left(\widehat{\Psi}(\mathbf{x}, t) f_t(\mathbf{x})\right) + \dfrac{1}{2}\beta_t \Delta \widehat{\Psi}(\mathbf{x}, t), \end{cases} \tag{6}$$

$$\text{s.t.} \quad \Psi(\mathbf{x}, 0)\widehat{\Psi}(\mathbf{x}, 0) = p_A(\mathbf{x}), \qquad \Psi(\mathbf{x}, 1)\widehat{\Psi}(\mathbf{x}, 1) = p_B(\mathbf{x}). \tag{7}$$

Compared with the diffusion dynamics in Eqs.(1, 2), the SB dynamics in Eqs.(4, 5) introduce an additional nonlinear forward drift $\beta_t \nabla_{\mathbf{x}} \log \Psi$, enabling transport between arbitrary boundary distributions. By Nelson's duality Nelson (1967), the two Schrödinger potentials satisfy $\Psi(\mathbf{x}, t)\widehat{\Psi}(\mathbf{x}, t) = q_t(\mathbf{x})$, where q_t denotes the marginal density of the bridge process. The backward potential gradient can be expressed as $\nabla_{\mathbf{x}} \log \widehat{\Psi} = \nabla_{\mathbf{x}} \log q_t - \nabla_{\mathbf{x}} \log \Psi$, and therefore is no longer the marginal score function. Using this relation in Eq.(5), we obtain $d\mathbf{X}_t = \left[f_t(\mathbf{X}_t) - \beta_t \left(\nabla_{\mathbf{x}} \log q_t(\mathbf{X}_t) - \nabla_{\mathbf{x}} \log \Psi(\mathbf{X}_t, t)\right)\right] dt + \sqrt{\beta_t}\, d\overline{W}_t$.

Tractable Schrödinger Bridge. The coupled boundary constraints in the Schrödinger system Eq.(7) make direct computation of the bridge dynamics intractable. Existing approaches typically rely on iterative procedures or specialized optimization schemes De Bortoli et al. (2021); Bernton et al. (2019), which can be computationally expensive in high-dimensional spaces. A more tractable perspective is to reformulate the Schrödinger dynamics within the score-based diffusion framework. Specifically, the unavailable nonlinear potential gradients in Eqs.(4) and (5) can be interpreted as score functions of auxiliary linear diffusion processes. The backward and forward potential gradients, $\nabla \log \widehat{\Psi}(\mathbf{X}_t, t)$ and $\nabla \log \Psi(\mathbf{X}_t, t)$, are governed by the following linear SDEs, Liu et al. (2023); Risken (1996) respectively:

$$d\mathbf{X}_t = f_t(\mathbf{X}_t)\, dt + \sqrt{\beta_t}\, dW_t, \qquad \mathbf{X}_0 \sim \widehat{\Psi}(\cdot, 0), \tag{8}$$

$$d\mathbf{X}_t = f_t(\mathbf{X}_t)\, dt + \sqrt{\beta_t}\, d\overline{W}_t, \qquad \mathbf{X}_1 \sim \Psi(\cdot, 1). \tag{9}$$

Although the auxiliary SDEs in Eqs.(8) and (9) are linear, their boundary distributions $\widehat{\Psi}(\cdot, 0)$ and $\Psi(\cdot, 1)$ are not directly available due to the coupled constraints in Eq.(7). To address this, a Dirac delta boundary condition Schwartz (1966) can be imposed at one endpoint, rendering the resulting conditional dynamics tractable. Specifically, for an individual initial sample $\mathbf{X}_0$, the initial distribution is represented as $p_A(\cdot) = \delta_{\mathbf{X}_0}(\cdot)$, which yields the following boundary conditions for the auxiliary linear SDEs:

$$\widehat{\Psi}(\cdot, 0) = \delta_{\mathbf{X}_0}(\cdot), \qquad \Psi(\cdot, 1) = \frac{p_B(\cdot)}{\widehat{\Psi}(\cdot, 1)}. \tag{10}$$

The former fixes the initial boundary at $\mathbf{X}_0$, independently of p_B. Consequently, the backward Schrödinger potential gradient $\nabla_{\mathbf{x}} \log \widehat{\Psi}(\mathbf{X}_t, t)$ corresponds to the conditional score of the auxiliary diffusion in Eq.(8), $\nabla_{\mathbf{x}} \log \widehat{\Psi}(\mathbf{X}_t, t) = \nabla_{\mathbf{x}} \log p^{\mathrm{aux}}(\mathbf{X}_t, t \mid \mathbf{X}_0)$, which provides a tractable target for standard score-based learning. To characterize the intermediate bridge states, we condition on paired endpoints $(\mathbf{X}_0, \mathbf{X}_1)$. In our setting, $\mathbf{X}_0$ corresponds to the clean peptide latent, while $\mathbf{X}_1$ represents the diffusion-generated intermediate state paired with it. When the base drift is set to $f_t = 0$, the resulting bridge admits a closed-form Gaussian posterior:

$$q_{\mathrm{B}}(\mathbf{X}_t \mid \mathbf{X}_0, \mathbf{X}_1) = \mathcal{N}\left(\mathbf{X}_t; \lambda_t^0 \mathbf{X}_0 + \lambda_t^1 \mathbf{X}_1, \rho_t^2 \mathbf{I}\right), \tag{11}$$

where λ_t^0 and λ_t^1 denote the time-dependent weights of the two endpoints, and ρ_t^2 determines the variance of the intermediate bridge state. Importantly, this analytic posterior is compatible with recursive DDPM-style sampling.

3.2 Unified Hybrid Diffusion–Bridge Dynamics

Variational AutoEncoder. Given a peptide graph $\mathcal{G}^p$ and its target binding-site graph $\mathcal{G}^b$, the encoder $\mathcal{E}_\phi$ maps the peptide–target complex into residue-level representations:

$$\mathbf{U}^p, \mathbf{U}^b = \mathcal{E}_\phi\left(\mathcal{G}^p, \mathcal{G}^b\right), \tag{12}$$

where $\mathbf{U}^p = \{(\mathbf{h}_i^p, \mathbf{x}_i^p)\}_{i=1}^{N_p}$ and $\mathbf{U}^b = \left\{\left(\mathbf{h}_i^b, \mathbf{x}_i^b\right)\right\}_{i=1}^{N_b}$. Here, $\mathbf{h}_i \in \mathbb{R}^d$ and $\mathbf{x}_i \in \mathbb{R}^3$ denote the invariant residue feature and equivariant coordinate representation, respectively. For each peptide residue, latent variables are sampled via the reparameterization trick Kingma & Welling (2014):

$$\mathbf{Z} = \{(\mathbf{z}_{h,i}, \mathbf{z}_{x,i})\}_{i=1}^{N_p}, \tag{13}$$

where $\mathbf{z}_{h,i}$ and $\mathbf{z}_{x,i}$ denote the semantic and coordinate latent variables, respectively. The decoder $\mathcal{D}_\xi$ reconstructs the peptide sequence and all-atom structure from the latent representation, conditioned on the binding-site features, $\widehat{\mathcal{G}}^p = \mathcal{D}_\xi\left(\mathbf{Z}, \mathbf{U}^b\right)$.

Hybrid Diffusion–Bridge Dynamics process. We formulate peptide generation as a hybrid stochastic process in the latent space, where diffusion and Schrödinger bridge govern different regimes along the generative trajectory. The resulting latent trajectory is expressed as:

$$\mathbf{Z}_T \xrightarrow{\text{Diffusion}} \mathbf{Z}_K \xrightarrow{\text{SB}} \mathbf{Z}_0. \tag{14}$$

Notably, the diffusion and Schrödinger bridge dynamics are coupled through the intermediate state $\mathbf{Z}_K$, which serves as their shared boundary. As shown in Eq. (11), the Schrödinger bridge is trained between paired endpoints, with each intermediate state conditioned on both $\mathbf{Z}_K$ and $\mathbf{Z}_0$. Since $\mathbf{Z}_0$ represents a clean, target-compatible peptide state, once diffusion establishes the coarse organization $\mathbf{Z}_K$, the remaining trajectory can be formulated as endpoint-conditioned stochastic transport toward $\mathbf{Z}_0$, progressively enforcing target-specific binding constraints. Their distinct stochastic dynamics are subsequently expressed under a common parameterization, enabling a unified learning formulation. We next establish a common representation for the two stochastic regimes. For score-based diffusion, the forward process is defined as:

$$q_{\mathrm{D}}(\mathbf{Z}_t \mid \mathbf{Z}_0) = \mathcal{N}\left(\mathbf{Z}_t; \sqrt{\bar{\alpha}_t}\mathbf{Z}_0, (1-\bar{\alpha}_t)\mathbf{I}\right), \tag{15}$$

$$\mathbf{Z}_t = a_t^{\mathrm{D}}\mathbf{Z}_0 + \sqrt{V_t^{\mathrm{D}}}\,\boldsymbol{\epsilon}, \qquad a_t^{\mathrm{D}} = \sqrt{\bar{\alpha}_t}, \qquad V_t^{\mathrm{D}} = 1-\bar{\alpha}_t. \tag{16}$$

where a_t^{D} and V_t^{D} denote the signal coefficient and noise variance at time t, respectively. At the shared boundary, the diffusion process yields the intermediate latent $\mathbf{Z}_K$. For the Schrödinger bridge dynamics, $\mathbf{Z}_0$ and $\mathbf{Z}_K$ define the paired endpoints. The closed-form Gaussian posterior can be reparameterized as:

$$q_{\mathrm{B}}\left(\mathbf{Z}_t \mid \mathbf{Z}_0, \mathbf{Z}_K\right) = \mathcal{N}\left(\mathbf{Z}_t; \lambda_t^0\mathbf{Z}_0 + \lambda_t^1\mathbf{Z}_K, \rho_t^2\mathbf{I}\right), \tag{17}$$

$$\mathbf{Z}_t = \lambda_t^0\mathbf{Z}_0 + \lambda_t^1\mathbf{Z}_K + \rho_t\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{18}$$

where $\lambda_t^0 = \frac{\bar{\sigma}_t^2}{\sigma_t^2+\bar{\sigma}_t^2}$, $\lambda_t^1 = \frac{\sigma_t^2}{\sigma_t^2+\bar{\sigma}_t^2}$, $\rho_t^2 = \frac{\sigma_t^2\bar{\sigma}_t^2}{\sigma_t^2+\bar{\sigma}_t^2}$. The accumulated variances are $\sigma_t^2 = \int_0^t \beta_s\,ds$, $\bar{\sigma}_t^2 = \int_t^1 \beta_s\,ds$, where β_s denotes the variance schedule of the Schrödinger bridge reference process. Substituting the diffusion boundary in Eq.(15) into Eq.(17) gives:

$$\mathbf{Z}_t^{\mathrm{B}} = a_t^{\mathrm{B}}\mathbf{Z}_0 + \sqrt{V_t^{\mathrm{B}}}\,\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{19}$$

where $a_t^{\mathrm{B}} = \lambda_t^0 + \lambda_t^1\sqrt{\bar{\alpha}_K}$, $V_t^{\mathrm{B}} = \left(\lambda_t^1\right)^2(1-\bar{\alpha}_K) + \rho_t^2$. Here, $\bar{\alpha}_K$ denotes the cumulative diffusion signal level associated with the intermediate boundary state $\mathbf{Z}_K$. Consequently, both stochastic regimes can be expressed within the same conditional Gaussian form:

$$q_r\left(\mathbf{Z}_t \mid \mathbf{Z}_0\right) = \mathcal{N}\left(\mathbf{Z}_t; a_t^{(r)}\mathbf{Z}_0, V_t^{(r)}\mathbf{I}\right), \qquad r \in \{\mathrm{D}, \mathrm{B}\}. \tag{20}$$

Here, $(a_t^{(\mathrm{D})}, V_t^{(\mathrm{D})})$ and $(a_t^{(\mathrm{B})}, V_t^{(\mathrm{B})})$ denote the diffusion and bridge coefficients, respectively. Despite their distinct transition dynamics, both regimes share a common conditional Gaussian parameterization conditioned on the clean latent $\mathbf{Z}_0$.

Algorithm 1 Training Procedure

Require: Peptide–target graphs $\mathcal{G}^p, \mathcal{G}^b$, $\mathbf{F}_\theta$
1: **repeat**
2: Obtain latent $\mathbf{Z}_0$ and $\mathbf{U}^b$ using $\mathcal{E}_\phi$
3: Select $r \in \{\mathrm{D}, \mathrm{B}\}$, t; set $(a_t^{(r)}, V_t^{(r)})$
4: Construct $\mathbf{Z}_t = a_t^{(r)}\mathbf{Z}_0 + \sqrt{V_t^{(r)}}\,\boldsymbol{\epsilon}$
5: Update $\mathbf{F}_\theta$ using Eq. (23)
6: **until** convergence

Algorithm 2 Sampling Procedure

Require: Target representation $\mathbf{U}^b$, trained $\mathbf{F}_\theta$
1: Sample $\mathbf{Z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: Predict $\widehat{\mathbf{Z}}_0$ using $\mathbf{F}_\theta$
4: Apply Eq. (21) if $t > K$; otherwise apply Eq. (22)
5: **end for**
6: **return** $\widehat{\mathcal{G}}^p = \mathcal{D}_\xi(\mathbf{Z}_0, \mathbf{U}^b)$

Hybrid Reverse Sampling. During inference, generation starts from a Gaussian latent $\mathbf{Z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The denoising network is used throughout the trajectory to estimate the clean latent $\widehat{\mathbf{Z}}_0$, while each stochastic regime follows its corresponding reverse transition. For the diffusion, the latent is progressively denoised from $t = T$ to $t = K$ according to the diffusion posterior:

$$p_{\mathrm{D}}\left(\mathbf{Z}_{t-1} \mid \mathbf{Z}_t, \widehat{\mathbf{Z}}_0\right) = \mathcal{N}\left(\mathbf{Z}_{t-1}; \widetilde{\boldsymbol{\mu}}_t^{\mathrm{D}}, \widetilde{\beta}_t \mathbf{I}\right), \qquad \widetilde{\boldsymbol{\mu}}_t^{\mathrm{D}} = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\widehat{\mathbf{Z}}_0 + \frac{\sqrt{\alpha_t}\,(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{Z}_t, \tag{21}$$

where $\widetilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. The resulting state $\mathbf{Z}_K$ initializes the reverse bridge dynamics. The latent is then recursively sampled toward the clean endpoint using the bridge posterior:

$$p_{\mathrm{B}}\left(\mathbf{Z}_{t-1}^{\mathrm{B}} \mid \mathbf{Z}_t^{\mathrm{B}}, \widehat{\mathbf{Z}}_0, \right) = \mathcal{N}\left(\mathbf{Z}_{t-1}^{\mathrm{B}}; \frac{\Delta\sigma_t^2}{\sigma_{t-1}^2 + \Delta\sigma_t^2}\widehat{\mathbf{Z}}_0 + \frac{\sigma_{t-1}^2}{\sigma_{t-1}^2 + \Delta\sigma_t^2}\mathbf{Z}_t^{\mathrm{B}}, \frac{\sigma_{t-1}^2\Delta\sigma_t^2}{\sigma_{t-1}^2 + \Delta\sigma_t^2}\mathbf{I}\right). \tag{22}$$

where $\Delta\sigma_t^2 = \sigma_t^2 - \sigma_{t-1}^2$ denotes the bridge variance accumulated between two consecutive reverse steps. The diffusion and bridge dynamics employ distinct recursive transition kernels but share the same network for estimating the clean latent $\widehat{\mathbf{Z}}_0$, enabling a unified reverse generation process under the shared target parameterization.

Unified Target Parameterization. Although the diffusion and bridge regimes share the Gaussian form in (20), their latent statistics vary with the corresponding signal and noise scales. To account for these regime-dependent statistics while retaining a shared denoising network, we introduce a preconditioned target parameterization:

$$\mathcal{L}_{\mathrm{gen}} = \mathbb{E}_{\mathbf{Z}_0, \mathbf{U}^b, t}\left[\left\|\mathbf{F}_\theta\left(\eta_t^{\mathrm{in}}\mathbf{Z}_t, t, \mathbf{U}^b\right) - \frac{\eta_t^{\mathrm{pred}}\mathbf{Z}_t - \mathbf{Z}_0}{\eta_t^{\mathrm{res}}}\right\|_2^2\right], \tag{23}$$

where η_t^{in}, η_t^{pred}, and η_t^{res} are preconditioning coefficients defined by the statistics of $(\mathbf{Z}_0, \mathbf{Z}_t)$ as:

$$\eta_t^{\mathrm{in}} = \frac{1}{\sqrt{\mathrm{Var}[\mathbf{Z}_t]}}, \qquad \eta_t^{\mathrm{pred}} = \frac{\mathrm{Cov}[\mathbf{Z}_0, \mathbf{Z}_t]}{\mathrm{Var}[\mathbf{Z}_t]}, \qquad \eta_t^{\mathrm{res}} = \sqrt{\mathrm{Var}[\mathbf{Z}_0] - \frac{\mathrm{Cov}[\mathbf{Z}_0, \mathbf{Z}_t]^2}{\mathrm{Var}[\mathbf{Z}_t]}}. \tag{24}$$

Here, η_t^{in} normalizes the input scale, η_t^{pred} provides the linear prediction of $\mathbf{Z}_0$ from $\mathbf{Z}_t$, and η_t^{res} normalizes the scale of the corresponding residual. The derivation of these coefficients is provided in the Appendix B.2. We use a 300-step generative trajectory, where ρ_D denotes the proportion of diffusion steps and p_B denotes the probability of sampling a training time step from the bridge segment. In the main experiments, we set $\rho_D = 0.3$, partitioning the 300-step trajectory into 90 diffusion steps and 210 bridge steps, with the dynamics switching at step 210 ($\mathbf{Z}_K = \mathbf{Z}_{210}$), as illustrated in Fig. 2. To strengthen the learning of target-directed transport toward target-compatible states, we set the bridge sampling probability to $p_B = 0.7$. Accordingly, 70% of training time steps are sampled from the bridge segment and 30% from the diffusion segment.

4 EXPERIMENTS

Dataset.

Our experiments are conducted on BioDB, which integrates protein–peptide complexes collected from the latest releases of Q-BioLip Wei et al. (2024) and PepBDB Wen et al. (2019). After removing redundant entries, we clustered receptor sequences using MMseqs2 Steinegger & Söding (2017) at a 40% sequence identity threshold. To prevent information leakage, complexes whose receptors belonged to the same cluster as any test-set receptor were excluded from the training and validation sets. The resulting dataset contains 11,058 training, 150 validation, and 147 test complexes.

Table 1: Quantitative comparison of different peptide generation methods on the LNR and BioDB datasets. All reported metrics are measured in % except RMSD_{C_α}. For each target, the compared methods generate 40 candidates for evaluation.

Dataset	Method	Recovery		Diversity		Fidelity		Physical Validity		Binding Quality				
		AAR↑	RMSD_{C_α} ↓	Seq↑	Struct↑	SSR↑	BSR↑	Clash_{in}↓	$\text{Clash}_{\text{out}}$↓	DockQ↑	SR↑	BNR_{Ros}↑	BNR_{MM}↑	BNR_{Evo}↑
LNR	PepFlow	34.01	4.1027	92.18	**56.40**	80.17	82.48	3.52	17.85	60.82	31.42	7.69	2.61	1.68
	PepGLAD	35.17	2.9728	89.76	39.65	84.47	82.10	3.63	3.98	68.25	54.78	13.98	6.20	3.80
	PepHAR	**36.29**	3.6343	16.67	31.48	79.79	81.08	10.37	20.08	61.64	27.20	8.60	3.48	2.83
	PepTir	34.59	2.8027	90.48	11.85	84.51	80.31	21.65	6.51	69.81	39.62	8.60	5.03	2.39
	PepBFN	29.28	4.7785	65.38	50.52	80.00	81.73	41.32	18.07	52.66	34.32	8.70	3.05	2.91
	UniMoMo	35.08	2.6672	91.58	22.53	92.78	93.62	**0.52**	1.46	75.15	82.77	33.33	9.70	9.29
	PepHyD	35.34	**1.8919**	**92.50**	13.44	**95.18**	**94.19**	0.78	**1.15**	**79.49**	**85.05**	**39.78**	**11.06**	**10.03**
BioDB	PepFlow	34.17	4.2162	**93.76**	**60.85**	79.89	81.78	4.85	15.03	57.58	40.97	16.98	4.52	3.63
	PepGLAD	34.07	3.5287	89.27	45.77	88.45	79.67	5.50	3.31	63.54	62.01	20.98	10.11	7.31
	PepHAR	31.37	2.8562	4.74	38.66	79.13	78.68	10.48	13.30	66.30	43.98	21.13	7.91	9.77
	PepTir	34.22	3.8093	90.47	17.01	89.08	77.78	28.57	5.19	61.67	45.09	13.99	9.34	7.39
	PepBFN	29.47	5.0094	66.67	54.30	80.60	80.81	43.98	15.05	52.34	38.13	12.95	5.43	3.72
	UniMoMo	33.54	2.9129	92.22	29.98	95.22	**93.09**	**0.48**	1.75	71.45	84.95	39.16	17.43	16.03
	PepHyD	**35.58**	**2.5581**	90.59	14.79	**96.34**	92.71	0.83	**0.68**	**75.34**	**87.01**	**40.56**	**18.07**	**16.36**

We additionally use 70k unlabeled protein fragment sequences from ProtFrag Kong et al. (2024) to train the variational autoencoder. To evaluate generalization to unseen targets, we further adopt the large non-redundant (LNR) benchmark Tsaban et al. (2022), which consists of 93 expert-curated protein–peptide complexes containing only canonical amino acids.

Baselines. We compare PepHyD with six representative target-conditioned peptide co-design methods: PepFlow Li et al. (2024b) adopts a multimodal flow-matching framework to jointly generate peptide sequences and structures by modeling residue types, backbone frames, and side-chain torsions in their respective spaces. PepGLAD Kong et al. (2024) encodes all-atom peptide residues into an equivariant latent representation and performs target-conditioned diffusion in the latent space. PepHAR Li et al. (2024a) first samples energetically favorable hotspot residues and then constructs complete peptides via autoregressive fragment extension. PepTri Nguyen et al. (2026) uses an latent diffusion framework with physics-informed, evolutionary, and mutual-information guidance for joint peptide sequence–structure generation. PepBFN Qian et al. (2026) uses Bayesian flow networks to jointly model residue identities and structural variables for all-atom peptide generation. UniMoMo Kong et al. (2025) uses geometric latent diffusion within a unified molecular representation for target-conditioned binder generation across peptides, antibodies, and small molecules.

Metrics. We evaluate generated peptides using the following metrics: **AAR** (Amino Acid Recovery) measures the fraction of residues matching the native peptide sequence. **RMSD_{C_α}** (root-mean-square deviation) measures the C_α deviation between generated and native peptides after target alignment. To assess diversity, we cluster generated candidates in sequence and structure spaces using thresholds of 40% sequence identity and 4 Å C_α RMSD. **Seq** and **Struct** denote the proportions of distinct sequence and structure clusters, respectively. **SSR** (Secondary Structure Ratio) measures the agreement between the DSSP-assigned secondary-structure labels of the generated and native peptides Kabsch & Sander (1983). **BSR** (Binding Site Rate) quantifies the overlap between their peptide–target binding sites. **Clash_{in}** and **$\text{Clash}_{\text{out}}$** measure steric clashes within the peptide and across the peptide–target interface, respectively, using a C_α–C_α distance threshold of 3.65 Å Ye et al. (2025). **DockQ** ranges from 0 to 1, with higher values indicating greater native-interface similarity Basu & Wallner (2016). We estimate binding energy ΔG using Rosetta Alford et al. (2017). **SR** denotes the fraction of candidates with $\Delta G < 0$, while **BNR_{Ros}**, **BNR_{MM}**, and **BNR_{Evo}** denote the fractions with lower binding energy than the native complex under Rosetta, MM-GBSA Kollman et al. (2000), and EvoEF2 Huang et al. (2020), respectively.

Results.

As shown in Table 1, PepHyD achieves strong overall performance on both LNR and BioDB. It attains the lowest RMSD_{C_α} and highest DockQ on both datasets, demonstrating improved structural fidelity and target-interface quality. PepHyD achieves the highest SR on both datasets, reaching 85.05% on LNR and 87.01% on BioDB, exceeding UniMoMo by 2.28 and 2.06 percentage points, respectively. It also attains the highest $\text{BNR}_{\text{Ros/MM/Evo}}$, indicating a higher proportion of generated peptides with stronger binding than the corresponding native complexes. Overall, these results

Table 2: Ablation study of different diffusion–bridge configurations on the BioDB dataset.

Setting	Recovery		Diversity		Fidelity		Physical Validity		Binding Quality				
	AAR↑	RMSD_{C_α} ↓	Seq↑	Struct↑	SSR↑	BSR↑	Clash_{in}↓	$\text{Clash}_{\text{out}}$↓	DockQ↑	SR↑	BNR_{Ros}↑	BNR_{MM}↑	BNR_{Evo}↑
$\rho_D = 0.1$	35.15	**2.5373**	**96.82**	21.50	**96.50**	92.72	**0.50**	0.90	73.41	85.84	37.76	17.29	15.00
$\rho_D = 0.5$	35.34	2.5853	96.66	18.31	95.18	92.84	1.35	1.32	74.63	85.75	38.46	17.79	16.17
$\rho_D = 0.7$	**35.72**	2.9372	96.54	27.36	94.20	**93.76**	2.36	1.73	71.04	80.49	37.76	16.24	14.52
Pure Bridge	35.60	2.9196	96.31	**30.77**	94.16	92.83	2.07	1.28	72.55	81.68	37.06	15.75	13.71
Pure Diffusion	34.11	2.6950	96.22	22.35	96.07	92.13	0.71	1.33	73.17	84.83	36.36	16.79	14.02
PepHyD (ρ_D = 0.3)	35.58	2.5581	90.59	14.79	96.34	92.71	0.83	**0.68**	**75.34**	**87.01**	**40.56**	**18.07**	**16.36**

Table 3: Binding-energy evaluation on three selected pockets with 40 generated peptides per method. Top-5 Mean is the mean of the five lowest ΔG values; #Strong counts candidates with $\Delta G < -5$.

Pocket	Method	SR (%) ↑	Top-5 Mean ΔG ↓	Best ΔG ↓	#Strong ↑
Pocket_1	UniMoMo	85.0	**−14.5**	**−18.8**	25
	PepHyD	**92.5**	−12.7	−15.6	**30**
Pocket_2	UniMoMo	70.0	−13.3	−15.3	16
	PepHyD	**82.5**	**−14.0**	**−18.4**	**18**
Pocket_3	UniMoMo	20.0	−8.9	−13.4	4
	PepHyD	**85.0**	**−15.8**	**−18.4**	**29**
Average	UniMoMo	58.3	−12.3	−15.8	15.0
	PepHyD	**86.7**	**−14.2**	**−17.5**	**25.7**

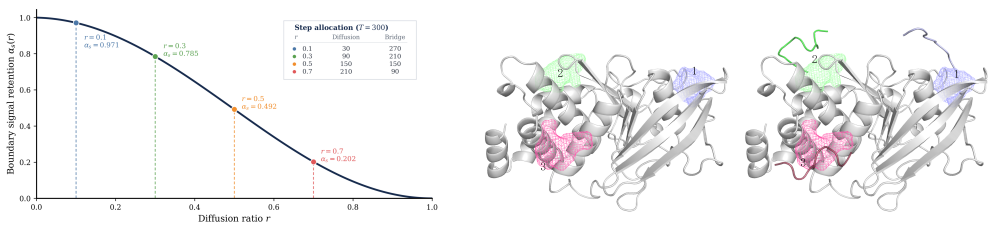
validate the effectiveness of the proposed hybrid generative dynamics for target-conditioned peptide co-design.

Ablation on Diffusion–Bridge Allocation.

To investigate the effect of diffusion–bridge allocation, we vary the diffusion proportion ρ_D over 0.1, 0.5, and 0.7, as illustrated in Fig. 2 (a), and further consider two extreme variants, Pure Bridge and Pure Diffusion, each using a single stochastic process throughout the trajectory. As shown in Table 2, different allocations lead to clear performance variations, highlighting the importance of coordinating the two dynamics. Small to moderate diffusion proportions generally maintain better structural fidelity and binding quality, whereas increasing ρ_D to 0.7 noticeably degrades RMSD_{C_α}, DockQ, and SR. Both pure variants also exhibit weaker binding performance than the default hybrid configuration. Overall, PepHyD with $\rho_D = 0.3$ achieves the strongest binding performance, supporting the effectiveness of combining diffusion-driven exploration with bridge-based target-directed transport.

4.1 Generalization to a New Therapeutic Target: PTP1B.

To further evaluate the generalization ability of PepHyD beyond the benchmark datasets, we conduct a target-specific study on human protein tyrosine phosphatase 1B (PTP1B) using the experimentally determined apo structure (PDB ID: 7KEY) Berman et al. (2000); Chirgadze et al. (2022). PTP1B is a key negative regulator of insulin and leptin signaling, and its dysregulation is strongly associated with type 2 diabetes and obesity, making it an important therapeutic target for metabolic disorders Zhang & Lee (2003). We select three candidate binding pockets on the PTP1B surface for pocket-conditioned peptide generation, as illustrated in Fig. 2 (b). Table 3 shows that PepHyD consistently achieves higher binding success rates than UniMoMo across all three pockets, with an average SR of 86.7%, outperforming UniMoMo by 28.4 percentage points. It also achieves better overall binding energies and generates more strongly binding candidates, with the most pronounced improvement on Pocket3, where SR increases from 20.0% to 85.0% and the number of



(a) Diffusion–Bridge Boundary Schedule

(b) Three binding pockets and generated peptides on PTP1B.

Figure 2: (a) Allocation across diffusion ratios ρ_D. (b) PTP1B binding pockets and representative PepHyD-generated peptides.

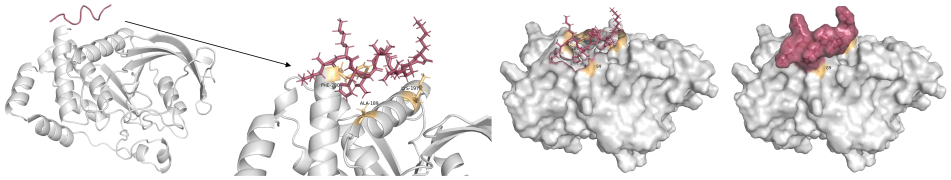


Figure 3: Example of a high-affinity binding conformation in PTP1B pocket 3 ($\Delta G = -18.4$).

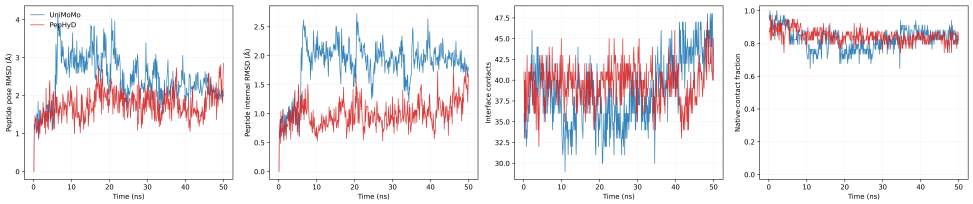


Figure 4: Molecular dynamics stability comparison of PepHyD and UniMoMo on target 4PIQ.

strong binders rises from 4 to 29. These results demonstrate the strong generalization ability of PepHyD, which can generate peptides with favorable binding properties across multiple pockets of an unseen therapeutically relevant target. A representative high-affinity binding conformation for PTP1B Pocket 3 is shown in Fig. 3.

4.2 Molecular Dynamics Analysis of Binding Stability

To further evaluate the dynamic stability of generated peptide–target complexes, we perform all-atom molecular dynamics (MD) simulations for PepHyD and UniMoMo on the 4PIQ target from the LNR dataset. For each method, we select the complex with the lowest Rosetta binding energy among 40 candidates. Detailed simulation settings are provided in Appendix A.1. We assess dynamic stability using the following metrics: Pose RMSD measures deviation from the initial binding configuration after receptor alignment; internal RMSD captures conformational changes within the peptide after peptide alignment; interface contacts quantify persistent peptide–target interactions; and the native-contact fraction measures the preservation of the initially generated interface.

As shown in Fig. 4, the two generated complexes exhibit distinct dynamic behaviors. PepHyD maintains lower pose and internal RMSDs throughout the trajectory. Its pose RMSD remains around 1.5–2.0 Å, whereas UniMoMo initially increases to approximately 3–4 Å. Similarly, the internal RMSD of PepHyD stays around 0.7–1.2 Å, compared with 1.8–2.2 Å for UniMoMo over most of the trajectory. These consistently lower RMSDs indicate reduced positional and conformational rearrangements for the PepHyD-generated peptide, suggesting a more stable binding configuration during simulation. For interface contacts, PepHyD maintains a more stable interaction pattern, whereas UniMoMo exhibits larger fluctuations and a pronounced mid-trajectory decrease. Similarly, PepHyD preserves a high native-contact fraction of around 0.8–0.9, while UniMoMo undergoes a larger decline during structural rearrangement. Overall, these results show that PepHyD better preserves the initial binding interface, undergoes less structural relaxation, and maintains a more stable binding geometry under explicit-solvent dynamics.

5 Conclusion

This paper presents PepHyD, a hybrid diffusion–Schrödinger dynamics framework for target-conditioned peptide sequence–structure co-design. By formulating generation as a unified stochastic trajectory with diffusion-driven exploration and Schrödinger bridge transport, PepHyD progressively transforms latent peptide representations toward target-compatible states while maintaining sequence–structure plausibility. Experiments on the LNR and BioDB benchmarks demonstrate state-of-the-art performance in structural fidelity and binding quality, while evaluation across multiple binding pockets of the unseen therapeutically relevant target PTP1B further confirms strong cross-target generalization. Molecular dynamics simulations additionally indicate enhanced binding stability and interface preservation. Nevertheless, the limited scale and target diversity of existing peptide–protein complex datasets may constrain generalization to broader target families. Future work will explore larger and more diverse datasets, as well as more generalizable generative dynamics for target-conditioned peptide design.

AI USE STATEMENT

In this work, we used large language models (LLMs) solely for language-related assistance, including proofreading, grammar checking, and improving the clarity and readability of the manuscript.

REFERENCES

Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Philip Bradley, Roland L. Dunbrack, Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, and Jeffrey J. Gray. The rosetta all-atom energy function for macromolecular modeling and design. *Journal of Chemical Theory and Computation*, 13(6):3031–3048, 2017. doi: 10.1021/acs.jctc.7b00125.

Sankar Basu and Björn Wallner. Dockq: A quality measure for protein–protein docking models. *PLOS ONE*, 11(8):e0161879, 2016. doi: 10.1371/journal.pone.0161879.

Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000. doi: 10.1093/nar/28.1.235.

Espen Bernton, Jeremy Heng, Arnaud Doucet, and Pierre E. Jacob. Schrödinger bridge samplers. *arXiv preprint arXiv:1912.13170*, 2019.

Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, Moises Lopez-Gonzalez, Alexander Mestre, Rio Watson, Sophia Vincoff, Lin Zhao, Jianli Wu, Audrey Stavrand, Mayumi Schaepers-Cheu, Tian Zi Wang, Divya Srijay, Connor Monticello, Pranay Vure, Rishab Pulugurta, Sarah Pertsemlidis, Kseniia Kholina, Shrey Goel, Matthew P. DeLisa, Jen-Tsan Ashley Chi, Ray Truant, Hector C. Aguilar, and Pranam Chatterjee. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, 44(6):1002–1010, 2026. doi: 10.1038/s41587-025-02761-2.

Sijie Chen, Tong Lin, Ruchira Basu, Jeremy Ritchey, Shen Wang, Yichuan Luo, Xingcan Li, Dehua Pei, Levent Burak Kara, and Xiaolin Cheng. Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations. *Nature Communications*, 15(1): 1611, 2024. doi: 10.1038/s41467-024-45766-2.

Zhaoqiang Chen, Xinben Zhang, Cheng Peng, Jinan Wang, Zhijian Xu, Kaixian Chen, Jiye Shi, and Weiliang Zhu. D3pockets: A method and web server for systematic analysis of protein pocket dynamics. *Journal of Chemical Information and Modeling*, 59(8):3353–3358, 2019. doi: 10.1021/acs.jcim.9b00332.

Yuri N. Chirgadze, Kevin P. Battaile, Ilya V. Likhachev, Nikolay K. Balabaev, Roni D. Gordon, Vladimir Romanov, Andres Lin, Robert Karisch, Robert Lam, Max Ruzanov, Evgeniy V. Brazhnikov, Emil F. Pai, Benjamin G. Neel, and Nickolay Y. Chirgadze. Signal transfer in human protein tyrosine phosphatase PTP1B from allosteric inhibitor P00058. *Journal of Biomolecular Structure and Dynamics*, 40(24):13823–13832, 2022. doi: 10.1080/07391102.2021.1994879.

Justas Dauparas, Ivan Anishchenko, Nathaniel Bennett, Hua Bai, Robert J. Ragotte, Lukas F. Milles, Basile I. M. 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. doi: 10.1126/science.add2187.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709, 2021.

Yuanqi Du, Arian R Jamasb, Jeff Guo, Tianfan Fu, Charles Harris, Yingheng Wang, Chenru Duan, Pietro Liò, Philippe Schwaller, and Tom L Blundell. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, 6(6):589–604, 2024.

Srinivasan Ekambaram and Nikolay V. Dokholyan. Peptide-based drug design using generative ai. *Chemical Communications*, 62(3):672–691, 2026. doi: 10.1039/D5CC04998A.

Caitlin L. Gare, Andrew M. White, and Lara R. Malins. From lead to market: chemical approaches to transform peptides into therapeutics. *Trends in Biochemical Sciences*, 50(6):467–480, 2025. doi: 10.1016/j.tibs.2025.01.009.

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

Xiaoqiang Huang, Robin Pearce, and Yang Zhang. EvoEF2: Accurate and fast energy function for computational protein design. *Bioinformatics*, 36(4):1135–1142, 2020. doi: 10.1093/bioinformatics/btz740.

Rui Jiao, Xiangzhe Kong, Ziyang Yu, Wenbing Huang, and Yang Liu. Equivariant pretrained transformer for unified geometric learning on multi-domain 3d molecules. In *ICLR 2024 Workshop on Generative and Experimental Perspectives for Biomolecular Design*, 2024.

Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12):2577–2637, 1983.

Beomsu Kim, Gihyun Kwon, Kwanyoung Kim, and Jong Chul Ye. Unpaired image-to-image translation via neural schrödinger bridge. In *International Conference on Learning Representations*, 2024.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Peter A. Kollman, Irina Massova, Carolina Reyes, Bernd Kuhn, Shuanghong Huo, Lillian Chong, Matthew Lee, Taisung Lee, Yong Duan, Wei Wang, Oreola Donini, Piotr Cieplak, Jaysharee Srinivasan, David A. Case, and Thomas E. Cheatham. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Accounts of Chemical Research*, 33(12):889–897, 2000. doi: 10.1021/ar000033j.

Xiangzhe Kong, Yinjun Jia, Wenbing Huang, and Yang Liu. Full-atom peptide design with geometric latent diffusion. *Advances in Neural Information Processing Systems*, 37:74808–74839, 2024.

Xiangzhe Kong, Zishen Zhang, Ziting Zhang, Rui Jiao, Jianzhu Ma, Wenbing Huang, Kai Liu, and Yang Liu. Unimomo: Unified generative modeling of 3d molecules for de novo binder design. *arXiv preprint arXiv:2503.19300*, 2025.

Jiahan Li, Tong Chen, Shitong Luo, Chaoran Cheng, Jiaqi Guan, Ruihan Guo, Sheng Wang, Ge Liu, Jian Peng, and Jianzhu Ma. Hotspot-driven peptide design via multi-fragment autoregressive extension. *arXiv preprint arXiv:2411.18463*, 2024a.

Jiahan Li, Chaoran Cheng, Zuofan Wu, Ruihan Guo, Shitong Luo, Zhizhou Ren, Jian Peng, and Jianzhu Ma. Full-atom peptide design based on multi-modal flow matching. *arXiv preprint arXiv:2406.00735*, 2024b.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A. Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22042–22062, 2023.

Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, and Liangqiong Qu. Residual denoising diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2773–2783, 2024.

Wenjie Liu, Yifan Zhu, Ying Zha, Qingshan Wu, Lei Jian, and Zhihao Liu. Rotation-and permutation-equivariant quantum graph neural network for 3d graph data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022.

Edward Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, 1967.

Ngoc-Quang Nguyen, Jaeyoon Jung, Seijung Kim, Sunkyu Kim, and Jaewoo Kang. Peptri: Tri-guided all-atom diffusion for peptide design via physics, evolution, and mutual information. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=yQlTgHo1um.

Achraf Oussidi and Azeddine Elhassouny. Deep generative models: Survey. In *2018 International conference on intelligent systems and computer vision (ISCV)*, pp. 1–8. IEEE, 2018.

Hao Qian, Shikui Tu, and Lei Xu. Full-atom peptide design via riemannian–euclidean bayesian flow networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(29):24918–24926, 2026. doi: 10.1609/aaai.v40i29.39679. URL https://ojs.aaai.org/index.php/AAAI/article/view/39679.

Hannes Risken. *The Fokker–Planck Equation: Methods of Solution and Applications*. Springer, 2 edition, 1996.

Laurent Schwartz. *Théorie des Distributions*. Hermann, Paris, 1966.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Linda Stein Gold, Kilian Eyerich, Joseph F. Merola, Joana Torres, Laura C. Coates, and Jessica R. Allegretti. Oral peptide therapeutics as an emerging treatment modality in immune-mediated inflammatory diseases: A narrative review. *Advances in Therapy*, 42(7):3158–3172, 2025. doi: 10.1007/s12325-025-03213-8.

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. doi: 10.1038/nbt.3988.

Tomer Tsaban, Julia K Varga, Orly Avraham, Ziv Ben-Aharon, Alisa Khramushin, and Ora Schueler-Furman. Harnessing protein folding neural networks for peptide–protein docking. *Nature communications*, 13(1):176, 2022.

Chentong Wang, Sarah Alamdari, Carles Domingo-Enrich, Ava P. Amini, and Kevin K. Yang. Toward deep learning sequence–structure co-generation for protein design. *Current Opinion in Structural Biology*, 91:103018, 2025a. doi: 10.1016/j.sbi.2025.103018.

Yuzhe Wang, Yibo Li, Jiaxiao Chen, and Luhua Lai. Modeling protein–ligand interactions for drug discovery in the era of deep learning. *Chemical Society Reviews*, 2025b.

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*, 620:1089–1100, 2023. doi: 10.1038/s41586-023-06415-8.

Hong Wei, Wenkai Wang, Zhenling Peng, and Jianyi Yang. Q-biolip: A comprehensive resource for quaternary structure-based protein–ligand interactions. *Genomics, Proteomics and Bioinformatics*, 22(1):qzae001, 2024.

Zeyu Wen, Jiahua He, Huanyu Tao, and Sheng-You Huang. Pepbdb: a comprehensive structural database of biological peptide–protein interactions. *Bioinformatics*, 35(1):175–177, 2019.

Xinyi Wu, Huitian Lin, Renren Bai, and Hongliang Duan. Deep learning for advancing peptide drug development: Tools and methods in structure prediction and design. *European Journal of Medicinal Chemistry*, 268:116262, 2024. doi: 10.1016/j.ejmech.2024.116262.

Wenjing Xiao, Wenjie Jiang, Zheng Chen, Yu Huang, Junyi Mao, Wei Zheng, Yonghe Hu, and Jianyou Shi. Advance in peptide-based drug development: delivery platforms, therapeutics and vaccines. *Signal Transduction and Targeted Therapy*, 10(1):74, 2025. doi: 10.1038/s41392-024-02107-5.

Qimeng Yang, Zhipeng Hu, Hongyu Jiang, Jialing Wang, Han Han, Wei Shi, and Hai Qian. Recent advances, strategies, and future perspectives of peptide-based drugs in clinical applications. *Chinese Journal of Natural Medicines*, 23(1):31–42, 2025. doi: 10.1016/S1875-5364(25)60800-4.

Fei Ye, Zaixiang Zheng, Dongyu Xue, Yuning Shen, Lihao Wang, Yiming Ma, Yan Wang, Xinyou Wang, Xiangxin Zhou, and Quanquan Gu. ProteinBench: A holistic evaluation of protein foundation models. In *International Conference on Learning Representations*, 2025.

Zhong-Yin Zhang and Seung-Yub Lee. PTP1B inhibitors as potential therapeutics in the treatment of type 2 diabetes and obesity. *Expert Opinion on Investigational Drugs*, 12(2):223–233, 2003. doi: 10.1517/13543784.12.2.223.

A SUPPLEMENTARY EXPERIMENTS

A.1 MOLECULAR DYNAMICS SIMULATION DETAILS AND VALIDATION ON PTP1B.

Setup. All simulations are conducted using OpenMM with the AMBER ff14SB force field and the TIP3P water model. Each complex is solvated in a periodic dodecahedral box with a minimum solute–box distance of 1.2 nm and an ionic strength of 0.15 M NaCl. Long-range electrostatic interactions are treated using the particle-mesh Ewald (PME) method with a 1.0 nm nonbonded cutoff. Bonds involving hydrogen atoms are constrained, allowing an integration time step of 2 fs. Each system is energy-minimized for up to 5,000 iterations and gradually heated to 298.15 K over 200 ps. The system is then equilibrated for 500 ps under the NVT ensemble, followed by 500 ps under the NPT ensemble at 1 atm with gradually released backbone restraints. Finally, a 50-ns unrestrained production simulation is performed.

Binding Pocket Detection Potential binding pockets on the target protein were identified using the D3Pockets web server (`https://www.d3pharma.com/D3Pocket/`) Chen et al. (2019). The protein structure was submitted in PDB format for pocket detection, and the predicted pockets were subsequently used for pocket-conditioned peptide generation.

Molecular Dynamics Metrics. We characterize the dynamic stability of the generated peptide–target complexes using four complementary metrics. **Peptide pose RMSD** measures the deviation of the peptide from its initially generated binding configuration after aligning the receptor, and therefore reflects changes in peptide position, orientation, and overall binding pose relative to the target; lower and more stable values indicate better preservation of the generated binding geometry. **Peptide internal RMSD** is computed after independently aligning the peptide and mainly captures conformational changes within the peptide itself, allowing rigid-body motion to be distinguished from internal structural rearrangement; lower values indicate better preservation of the peptide conformation. **Interface contacts** quantify the number of direct peptide–target contacts maintained during the simulation and reflect the persistence of physical association at the binding interface; a relatively stable contact number suggests sustained peptide–target interactions, whereas a pronounced decrease indicates interface weakening or disruption. **Native-contact fraction** measures the proportion of contacts from the initially generated interface that remain preserved over time, with values closer to 1 indicating stronger retention of the original interaction pattern.

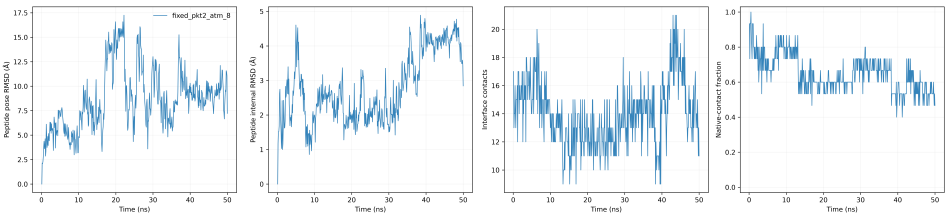


Figure 5: Molecular dynamics analysis of the generated peptide–PTP1B complex in pocket 3 over a 50-ns simulation.

Figure 3 shows a representative high-affinity peptide bound to PTP1B Pocket 3. We further perform molecular dynamics simulations to evaluate the dynamic stability of this complex under explicit-solvent conditions. The peptide exhibits some variation in its binding pose during the simulation, while its internal RMSD remains considerably lower over most of the trajectory, indicating that the dominant change arises from repositioning at the target surface rather than complete loss of peptide conformation. Meanwhile, the peptide maintains approximately 10–20 interface contacts throughout the trajectory, and the native-contact fraction remains mostly within 0.5–0.8, indicating that a substantial portion of the initial peptide–target interactions is retained despite the change in binding geometry. The larger pose deviation may partly reflect uncertainty in the initial binding-site specification: unlike benchmark complexes with experimentally resolved peptide–target interfaces, the PTP1B pockets are computationally identified from the apo structure using D3Pocket and therefore provide only approximate candidate binding regions rather than experimentally determined peptide poses. Overall, the simulation suggests that the generated peptide remains associated with the predicted pocket while undergoing substantial structural relaxation under explicit-solvent dynamics.

A.2 Effect of Candidate Generation Budget.

Table 4: Ablation study on the number of generated samples on the BioDB dataset.

Samples	Recovery		Diversity		Fidelity		Physical Validity		Binding Quality				
	AAR↑	RMSD_{C_α} ↓	Seq↑	Struct↑	SSR↑	BSR↑	Clash_{in}↓	$\text{Clash}_{\text{out}}$↓	DockQ↑	SR↑	BNR_{Ros}↑	BNR_{MM}↑	BNR_{Evo}↑
$N = 10$	28.88	3.0476	**99.37**	**35.59**	91.42	89.31	**0.69**	0.85	69.13	**87.83**	27.97	16.06	16.29
$N = 20$	31.70	2.9212	98.36	23.39	94.31	91.54	0.79	0.82	72.50	86.22	32.87	17.95	15.51
$N = 30$	33.25	2.6423	97.32	17.65	95.82	92.65	0.73	0.81	74.17	86.32	37.06	17.73	15.92
$N = 60$	36.48	2.5448	95.36	11.55	97.02	94.33	0.79	0.85	76.34	86.25	42.66	17.63	15.53
$N = 100$	**39.48**	**2.3299**	92.66	8.66	**97.80**	**95.26**	0.77	0.84	**77.98**	86.46	**46.15**	17.87	15.84
PepHyD (40)	35.58	2.5581	90.59	14.79	96.34	92.71	0.83	**0.68**	75.34	87.01	40.56	**18.07**	**16.36**

To evaluate the effect of the sampling budget, we vary the number of peptides generated for each target as $N \in 10, 20, 30, 60, 100$ on BioDB, with $N = 40$ used in the main experiments. As shown in Table 4, As N increases, recovery, structural fidelity, and BNR_{Ros} generally improve, as a larger candidate pool increases the chance of obtaining high-quality peptides. Nevertheless, PepHyD already maintains a stable SR of around 86–88% with as few as 10–30 samples, while BNR_{MM} and BNR_{Evo} remain comparable across different sampling budgets. Importantly, these results indicate that the strong binding performance of PepHyD does not rely primarily on large-scale candidate generation or extensive post-generation screening. PepHyD remains effective even with a limited sampling budget.

A.3 Ablation on Bridge Sampling Probability.

Table 5: Ablation study on the bridge sampling probability on the BioDB dataset.

Setting	Recovery		Diversity		Fidelity		Physical Validity		Binding Quality				
	AAR↑	RMSD_{C_α} ↓	Seq↑	Struct↑	SSR↑	BSR↑	Clash_{in}↓	$\text{Clash}_{\text{out}}$↓	DockQ↑	SR↑	BNR_{Ros}↑	BNR_{MM}↑	BNR_{Evo}↑
$p_B = 0.3$	**36.12**	2.7313	**96.57**	**28.48**	94.20	92.29	1.28	1.24	72.65	85.48	34.27	16.26	14.71
$p_B = 0.5$	35.25	2.7826	96.38	21.88	94.73	93.04	1.06	1.17	72.71	87.22	37.76	16.27	14.97
$p_B = 0.9$	34.29	2.6329	94.45	19.48	**96.73**	**93.26**	0.93	0.81	74.22	**87.72**	39.16	17.52	15.30
PepHyD ($p_B = 0.7$)	35.58	**2.5581**	90.59	14.79	96.34	92.71	**0.83**	**0.68**	**75.34**	87.01	**40.56**	**18.07**	**16.36**

To investigate the effect of bridge sampling during training, we vary the bridge sampling probability as $p_B \in 0.3, 0.5, 0.7, 0.9$, with $p_B = 0.7$ used as the default setting. As shown in Table 5, increasing p_B generally improves structural fidelity and binding quality, indicating the benefit of sufficiently emphasizing bridge dynamics during training. However, assigning an excessively high probability to the bridge regime does not provide further overall gains. Although $p_B = 0.9$ achieves a slightly higher SR, the default $p_B = 0.7$ yields better RMSD_{C_α}, DockQ, physical validity, and BNR across all three energy functions. Overall, these results suggest that a balanced training allocation between diffusion and bridge regimes is important, with $p_B = 0.7$ providing the most favorable overall performance.

B Methods Details

B.1 Graph Representation and Model Architecture

We model the binding site and peptide as residue-level block graphs, represented by $\mathcal{G}^b = (\mathcal{V}^b, \mathcal{E}^b)$ and $\mathcal{G}^p = (\mathcal{V}^p, \mathcal{E}^p)$, respectively. In both graphs, each node $i \in \mathcal{V}$ represents an amino acid residue and is modeled as an atomistic block $(A_i, \mathbf{X}_i)$. Specifically, $A_i \in \mathbb{Z}^{n_i}$ records the atom types of the n_i atoms belonging to residue i, while $\mathbf{X}_i \in \mathbb{R}^{n_i \times 3}$ contains their corresponding three-dimensional coordinates. The edge set $\mathcal{E}$ encodes the chemical connectivity among atoms, including both intra-residue and inter-residue bonds. We use $N_b = |\mathcal{V}^b|$ and $N_p = |\mathcal{V}^p|$ to denote the numbers of residues in the target binding site and peptide, respectively. For each protein–peptide complex, the binding site is defined as the set of target residues whose C_β atoms lie within 10 Å of any peptide atom, consistent with prior target-conditioned peptide generation settings Kong et al. (2024). This

definition retains sufficient local interaction context while avoiding an overly narrow binding pocket. The backbone of PepHyD is implemented using an Equivariant Transformer (ET) Jiao et al. (2024), which maintains equivariance under global rotations and translations Liu et al. (2025).

B.2 VARIATIONAL AUTOENCODER

We employ a variational autoencoder (VAE) to learn a continuous latent representation that jointly captures peptide sequence and structural information. Given a peptide graph $\mathcal{G}^p$ and its target binding-site graph $\mathcal{G}^b$, the encoder $\mathcal{E}_\phi$ first produces residue-level peptide and target representations:

$$\mathbf{U}^p, \mathbf{U}^b = \mathcal{E}_\phi\left(\mathcal{G}^p, \mathcal{G}^b\right), \tag{25}$$

where $\mathbf{U}^p = \{(\mathbf{h}_i^p, \mathbf{x}_i^p)\}_{i=1}^{N_p}$. Here, $\mathbf{h}_i^p$ represents the invariant residue feature and $\mathbf{x}_i^p$ denotes its equivariant coordinate representation.

Latent reparameterization. For each peptide residue, the encoder parameterizes an invariant semantic latent $\mathbf{z}_{h,i}$ and an equivariant coordinate latent $\mathbf{z}_{x,i}$. Their posterior distributions are represented by Gaussian distributions,

$$q_\phi(\mathbf{z}_{h,i}) = \mathcal{N}\left(\boldsymbol{\mu}_{h,i}, \mathrm{diag}(\boldsymbol{\sigma}_{h,i}^2)\right), \qquad q_\phi(\mathbf{z}_{x,i}) = \mathcal{N}\left(\boldsymbol{\mu}_{x,i}, \sigma_{x,i}^2\mathbf{I}\right). \tag{26}$$

The latent variables are sampled using the standard reparameterization trick:

$$\mathbf{z}_{h,i} = \boldsymbol{\mu}_{h,i} + \boldsymbol{\sigma}_{h,i} \odot \boldsymbol{\epsilon}_{h,i}, \qquad \mathbf{z}_{x,i} = \boldsymbol{\mu}_{x,i} + \sigma_{x,i}\boldsymbol{\epsilon}_{x,i}, \tag{27}$$

where $\boldsymbol{\epsilon}_{h,i}$ and $\boldsymbol{\epsilon}_{x,i}$ are sampled from standard Gaussian distributions. The resulting peptide latent representation is

$$\mathbf{Z}^p = \{(\mathbf{z}_{h,i}, \mathbf{z}_{x,i})\}_{i=1}^{N_p}. \tag{28}$$

Sequence and structure decoding. The decoder $\mathcal{D}_\xi$ reconstructs the peptide conditioned on both the peptide latent representation $\mathbf{Z}^p$ and the target representation $\mathbf{U}^b$. For sequence reconstruction, the semantic and coordinate latents are jointly decoded to predict the amino-acid identity of each peptide residue:

$$\widehat{s}_i^p = \mathcal{D}_\xi^{\mathrm{seq}}\left(\mathbf{Z}^p, \mathbf{U}^b\right)_i, \tag{29}$$

where $\widehat{s}_i^p$ denotes the predicted residue-type distribution for residue i.

For structural reconstruction, atom coordinates are initialized around the corresponding residue coordinate latent $\mathbf{z}_{x,i}$. Given an intermediate noisy structure $\mathbf{X}_t$, the structural decoder predicts the atomic displacement toward the native peptide structure, while reconstructing the complete peptide graph:

$$\widehat{\mathbf{V}} = \mathcal{D}_\xi^{\mathrm{str}}(\mathbf{Z}^p, \mathbf{X}_t, t, \mathbf{U}^b), \qquad \widehat{\mathcal{G}}^p = \mathcal{D}_\xi(\mathbf{Z}^p, \mathbf{U}^b), \tag{30}$$

where $\widehat{\mathbf{V}}$ denotes the predicted atomic displacement, and $\widehat{\mathcal{G}}^p$ contains the reconstructed residue identities and all-atom coordinates.

Training objective. The VAE is trained with latent regularization together with sequence and structure reconstruction objectives.

For the semantic and coordinate latent variables, we employ KL regularization:

$$\mathcal{L}_{\mathrm{KL}}^h = \frac{1}{N_p}\sum_{i=1}^{N_p} D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}_{h,i}) \| \mathcal{N}(\mathbf{0}, \mathbf{I})\right), \tag{31}$$

$$\mathcal{L}_{\mathrm{KL}}^x = \frac{1}{N_p}\sum_{i=1}^{N_p} D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}_{x,i}) \| \mathcal{N}(\mathbf{c}_i, \mathbf{I})\right), \tag{32}$$

where $\mathbf{c}_i$ denotes the geometric center of residue i.

For sequence reconstruction, the decoder predicts the amino-acid identity of each peptide residue. We use the cross-entropy loss

$$\mathcal{L}_{\mathrm{seq}} = -\frac{1}{N_p}\sum_{i=1}^{N_p} \log \widehat{p}_i(s_i), \tag{33}$$

where s_i is the ground-truth residue type and $\widehat{p}_i(s_i)$ is its predicted probability.

For structural reconstruction, the decoder predicts the atomic displacement toward the ground-truth peptide structure:

$$\mathcal{L}_{\text{coord}} = \frac{1}{N_a} \sum_{a=1}^{N_a} \|\widehat{\mathbf{v}}_a - \mathbf{v}_a\|_2^2 , \tag{34}$$

where N_a is the number of peptide atoms, and $\mathbf{v}_a$ and $\widehat{\mathbf{v}}_a$ denote the ground-truth and predicted atomic displacements, respectively.

To further preserve local peptide geometry, we constrain pairwise atomic distances:

$$\mathcal{L}_{\text{local}} = \frac{1}{|\mathcal{P}|} \sum_{(a,b)\in\mathcal{P}} \text{SmoothL1}\left(\widehat{d}_{ab}, d_{ab}\right), \tag{35}$$

where $\mathcal{P}$ denotes the selected local atom pairs, and d_{ab} and $\widehat{d}_{ab}$ are their ground-truth and reconstructed distances, respectively.

The complete VAE objective is

$$\boxed{\mathcal{L}_{\text{VAE}} = 0.6\,\mathcal{L}_{\text{KL}}^{h} + 0.8\,\mathcal{L}_{\text{KL}}^{x} + \mathcal{L}_{\text{seq}} + \mathcal{L}_{\text{coord}} + 0.5\,\mathcal{L}_{\text{local}}} \tag{36}$$

where the loss weights are set following prior work Kong et al. (2024; 2025).

B.3 Additional Details of Hybrid Diffusion–Bridge Dynamics

Unified Gaussian representation. As described in the main text, the diffusion and bridge regimes are coupled through the intermediate latent $\mathbf{Z}_K$. The diffusion marginal at the switching point is

$$\mathbf{Z}_K = \sqrt{\bar{\alpha}_K}\mathbf{Z}_0 + \sqrt{1-\bar{\alpha}_K}\boldsymbol{\epsilon}_K, \qquad \boldsymbol{\epsilon}_K \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{37}$$

For the bridge regime, the intermediate state is given by

$$\mathbf{Z}_t^B = \lambda_t^0 \mathbf{Z}_0 + \lambda_t^1 \mathbf{Z}_K + \rho_t \boldsymbol{\epsilon}_B, \qquad \boldsymbol{\epsilon}_B \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{38}$$

Substituting $\mathbf{Z}_K$ gives

$$\mathbf{Z}_t^B = \left(\lambda_t^0 + \lambda_t^1\sqrt{\bar{\alpha}_K}\right)\mathbf{Z}_0 + \lambda_t^1\sqrt{1-\bar{\alpha}_K}\,\boldsymbol{\epsilon}_K + \rho_t\boldsymbol{\epsilon}_B. \tag{39}$$

Since $\boldsymbol{\epsilon}_K$ and $\boldsymbol{\epsilon}_B$ are independent Gaussian variables, the bridge marginal can be rewritten as

$$\mathbf{Z}_t^B = a_t^B \mathbf{Z}_0 + \sqrt{V_t^B}\boldsymbol{\epsilon}, \tag{40}$$

where

$$a_t^B = \lambda_t^0 + \lambda_t^1\sqrt{\bar{\alpha}_K}, \qquad V_t^B = (\lambda_t^1)^2(1-\bar{\alpha}_K) + \rho_t^2. \tag{41}$$

Together with the diffusion coefficients $a_t^D = \sqrt{\bar{\alpha}_t}$ and $V_t^D = 1-\bar{\alpha}_t$, both regimes therefore admit the unified form

$$\mathbf{Z}_t = a_t^{(r)}\mathbf{Z}_0 + \sqrt{V_t^{(r)}}\boldsymbol{\epsilon}, \qquad r \in \{\text{D}, \text{B}\}. \tag{42}$$

Derivation of the unified target parameterization. Given the unified latent perturbation

$$\mathbf{Z}_t = a_t^{(r)}\mathbf{Z}_0 + \sqrt{V_t^{(r)}}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{43}$$

we derive a common preconditioning scheme for both stochastic regimes. For clarity, we omit the regime superscript r below. The coefficient η_t^{in} normalizes the perturbed latent, η_t^{pred} provides a linear estimate of the clean latent, and η_t^{res} normalizes the remaining prediction residual.

Since $\boldsymbol{\epsilon}$ is independent of $\mathbf{Z}_0$,

$$\text{Var}[\mathbf{Z}_t] = a_t^2\,\text{Var}[\mathbf{Z}_0] + V_t, \qquad \text{Cov}[\mathbf{Z}_0, \mathbf{Z}_t] = a_t\,\text{Var}[\mathbf{Z}_0]. \tag{44}$$

To maintain a normalized network input, we require $\text{Var}[\eta_t^{\text{in}}\mathbf{Z}_t] = 1$, which gives

$$\eta_t^{\text{in}} = \frac{1}{\sqrt{a_t^2\,\text{Var}[\mathbf{Z}_0] + V_t}}. \tag{45}$$

We next consider a linear estimate $\eta_t^{\text{pred}}\mathbf{Z}_t$ of the clean latent. The optimal coefficient is obtained by minimizing the mean-squared prediction error:

$$\eta_t^{\text{pred}} = \arg\min_{\eta} \mathbb{E}\left[\|\mathbf{Z}_0 - \eta\mathbf{Z}_t\|_2^2\right] = \frac{\text{Cov}[\mathbf{Z}_0, \mathbf{Z}_t]}{\text{Var}[\mathbf{Z}_t]}. \tag{46}$$

Substituting the moments above yields

$$\eta_t^{\text{pred}} = \frac{a_t \,\text{Var}[\mathbf{Z}_0]}{a_t^2 \,\text{Var}[\mathbf{Z}_0] + V_t}. \tag{47}$$

The variance of the remaining residual $\mathbf{Z}_0 - \eta_t^{\text{pred}}\mathbf{Z}_t$ is

$$\text{Var}\left[\mathbf{Z}_0 - \eta_t^{\text{pred}}\mathbf{Z}_t\right] = \text{Var}[\mathbf{Z}_0] - \frac{\text{Cov}[\mathbf{Z}_0, \mathbf{Z}_t]^2}{\text{Var}[\mathbf{Z}_t]}. \tag{48}$$

Therefore, the residual normalization coefficient is

$$\eta_t^{\text{res}} = \sqrt{\text{Var}[\mathbf{Z}_0] - \frac{a_t^2 \,\text{Var}[\mathbf{Z}_0]^2}{a_t^2 \,\text{Var}[\mathbf{Z}_0] + V_t}}. \tag{49}$$

These coefficients normalize the input and prediction residual across different noise levels, allowing the diffusion and bridge regimes to share the same denoising target despite having different $(a_t^{(r)}, V_t^{(r)})$ schedules.

C Implementation Details

C.1 Training

PepHyD is trained in two stages. We first train the VAE to learn a continuous latent representation of peptide sequence and structure. The VAE is optimized with AdamW using an initial learning rate of 1×10^{-4} and gradient clipping at 1.0. We train the VAE for up to 250 epochs and apply a ReduceLROnPlateau scheduler with a decay factor of 0.8, a patience of five epochs, and a minimum learning rate of 5×10^{-6}. After VAE training, its parameters are frozen and the peptide–target complexes are encoded into the latent space for training the hybrid diffusion–bridge dynamics. The shared denoising network is optimized with AdamW using an initial learning rate of 1×10^{-4} for up to 500 epochs. We use a 10-epoch linear warm-up followed by cosine learning rate decay, with a minimum learning rate of 5×10^{-6}. Early stopping is applied with a patience of 10 epochs.

C.2 Datasets and Preprocessing

We conduct experiments on BioDB and LNR, with ProtFrag additionally used for VAE pretraining. The dataset statistics are reported in Table 6.

BioDB is assembled from peptide–protein complexes collected from Q-BioLip Wei et al. (2024) and PepBDB Wen et al. (2019). After merging the two sources, duplicated complexes are removed and only peptides with lengths between 3 and 25 residues are retained. To reduce redundancy between data splits, receptor sequences are grouped using MMseqs2 Steinegger & Söding (2017) with a 40% sequence-identity threshold. All complexes associated with receptors from the same cluster are kept within the same split. This procedure results in 11,058 training complexes, 150 validation complexes, and 147 test complexes. For each peptide–protein complex, we extract a local target binding site rather than using the entire receptor structure. Specifically, a target residue is included if its C_β atom is within $10\,\text{Å}$ of any peptide atom. The resulting local environment is used as the target graph for conditional peptide generation. In addition to BioDB, approximately 70,000 protein fragments from ProtFrag Kong et al. (2024) are used during VAE training to improve the learning of peptide sequence–structure representations. We further evaluate cross-target generalization on the large non-redundant (LNR) benchmark Tsaban et al. (2022), which contains 93 curated peptide–protein complexes composed of canonical amino acids. LNR is used exclusively for evaluation and is not involved in model training or hyperparameter selection.

Table 6: Dataset statistics used for training and evaluation.

Dataset	Split	# Complexes	# Clusters
BioDB	Training	11,058	1,531
	Validation	150	150
	Test	147	147
LNR	Test	93	93

C.3 Baseline Implementations

To ensure a consistent comparison, PepFlow, PepGLAD, PepHAR, PepTri, and PepBFN are trained on BioDB using their publicly available implementations. We use the same BioDB training, validation, and test splits for all methods, while following the default network configurations and optimization settings provided by the corresponding official codebases. For UniMoMo, we use the officially released pretrained checkpoint because a peptide-specific checkpoint is not provided. This checkpoint was trained jointly on approximately 184K molecular samples spanning multiple binder modalities.At inference time, UniMoMo is evaluated using the same target binding sites, peptide lengths, and number of generated candidates as PepHyD.

C.4 Evaluation Metrics

We evaluate generated peptides from five aspects: recovery, diversity, structural fidelity, physical validity, and binding quality. Let $\mathcal{T}$ denote the set of target binding sites and $\mathcal{C}_k$ the set of candidates generated for target k. Unless otherwise stated, ratio-based metrics and DockQ are reported in percentage, while RMSD_{C_α} is reported in Å.

Recovery and structural fidelity. Amino Acid Recovery (AAR) measures the sequence agreement between the generated peptide and its reference sequence after alignment. RMSD_{C_α} measures the structural deviation between matched peptide C_α atoms after aligning the corresponding target structures. For each target, we retain the candidate with the highest AAR and the candidate with the lowest RMSD_{C_α}. The reported metrics are

$$\mathrm{AAR} = \frac{1}{|\mathcal{T}|} \sum_{k \in \mathcal{T}} \max_{j \in \mathcal{C}_k} \mathrm{AAR}_{k,j}, \tag{50}$$

$$\mathrm{RMSD}_{C_\alpha} = \mathrm{Median}_{k \in \mathcal{T}} \left[\min_{j \in \mathcal{C}_k} \mathrm{RMSD}_{C_\alpha}^{k,j} \right]. \tag{51}$$

Diversity. We measure diversity independently in sequence and structure spaces. For each target, generated peptides are clustered using a 40% sequence identity threshold, while structural clustering is performed using a pairwise C_α RMSD threshold of 4 Å. The sequence and structure diversity scores are defined as

$$\mathrm{Seq}_k = \frac{N_{\mathrm{seq}}^k}{|\mathcal{C}_k|}, \qquad \mathrm{Struct}_k = \frac{N_{\mathrm{struct}}^k}{|\mathcal{C}_k|}, \tag{52}$$

where N_{seq}^k and N_{struct}^k denote the numbers of sequence and structure clusters for target k, respectively. Both metrics are first computed for each target and then averaged across all targets.

Secondary-structure fidelity. Secondary Structure Ratio (SSR) evaluates the agreement between the DSSP assignments of generated and reference peptides. DSSP states are mapped into three classes: $\{\mathrm{H}, \mathrm{G}, \mathrm{I}\} \to \mathrm{H}$, $\{\mathrm{E}, \mathrm{B}\} \to \mathrm{E}$, and all remaining states to C. For candidate j of target k,

$$\mathrm{SSR}_{k,j} = \frac{1}{L} \sum_{r=1}^{L} \mathbb{I}\left[s_r^{\mathrm{gen}} = s_r^{\mathrm{ref}} \right], \tag{53}$$

where L denotes the aligned sequence length and s_r^{gen} and s_r^{ref} are the corresponding secondary-structure labels. We report the mean target-wise best SSR:

$$\mathrm{SSR} = \frac{1}{|\mathcal{T}|} \sum_{k \in \mathcal{T}} \max_{j \in \mathcal{C}_k} \mathrm{SSR}_{k,j}. \tag{54}$$

Binding-site fidelity. Binding Site Rate (BSR) evaluates how well the generated peptide preserves the reference binding region on the target protein. A target residue is considered part of the binding site if its C_β atom lies within 6 Å of the peptide, with C_α used when C_β is unavailable. Let $\mathcal{S}_k^{\mathrm{ref}}$ and $\mathcal{S}_{k,j}^{\mathrm{gen}}$ denote the binding-site residue sets of the reference and generated complexes, respectively. We define

$$\mathrm{BSR}_{k,j} = \frac{\left|\mathcal{S}_k^{\mathrm{ref}} \cap \mathcal{S}_{k,j}^{\mathrm{gen}}\right|}{\left|\mathcal{S}_k^{\mathrm{ref}}\right|}, \tag{55}$$

and report the mean target-wise best score:

$$\mathrm{BSR} = \frac{1}{|\mathcal{T}|} \sum_{k \in \mathcal{T}} \max_{j \in \mathcal{C}_k} \mathrm{BSR}_{k,j} \,. \tag{56}$$

Physical validity. We assess steric validity using intra-peptide and peptide–target clash rates. $\mathrm{Clash}_{\mathrm{in}}$ measures the fraction of peptide residues involved in intrapeptide clashes, where two non-adjacent C_α atoms are considered clashing if their distance is below 3.6574 Å. $\mathrm{Clash}_{\mathrm{out}}$ measures the fraction of peptide residues whose C_α atoms are within 3.6574 Å of any target C_α atom. Both metrics are averaged over generated candidates for each target and subsequently averaged across targets.

Binding quality. DockQ Basu & Wallner (2016) evaluates the similarity of the generated protein–peptide interface to the reference complex. For each target, we retain the candidate with the highest DockQ and report

$$\mathrm{DockQ} = \frac{1}{|\mathcal{T}|} \sum_{k \in \mathcal{T}} \max_{j \in \mathcal{C}_k} \mathrm{DockQ}_{k,j} \,. \tag{57}$$

We further evaluate energetic binding quality using Rosetta, MM-GBSA, and EvoEF2. For Rosetta, the binding energy is computed as

$$\Delta G = E_{\mathrm{bound}} - E_{\mathrm{separated}}, \tag{58}$$

where lower ΔG indicates more favorable predicted binding. The Success Rate (SR) is defined as the fraction of generated candidates satisfying $\Delta G < 0$.

To compare generated complexes against their corresponding native complexes, we additionally report the Beat-Native Ratio (BNR). For a scoring function f, it is defined as

$$\mathrm{BNR}_f = \frac{\sum_{k \in \mathcal{T}} \sum_{j \in \mathcal{C}_k^f} \mathbb{I}\left[E_f^{k,j} < E_f^{k,\mathrm{GT}}\right]}{\sum_{k \in \mathcal{T}} |\mathcal{C}_k^f|}, \tag{59}$$

where $E_f^{k,j}$ and $E_f^{k,\mathrm{GT}}$ denote the scores of generated candidate j and the corresponding native complex, respectively, and $\mathcal{C}_k^f$ denotes the candidates successfully evaluated by scoring function f. We report $\mathrm{BNR}_{\mathrm{Ros}}$, $\mathrm{BNR}_{\mathrm{MM}}$, and $\mathrm{BNR}_{\mathrm{Evo}}$ using Rosetta, MM-GBSA Kollman et al. (2000), and EvoEF2 Huang et al. (2020), respectively. Higher BNR values indicate that a larger fraction of generated peptides achieve more favorable predicted binding energies than their native counterparts.

C.5 Hyperparameter Settings

Table 7 summarizes the main hyperparameters of PepHyD, including the VAE, hybrid diffusion–bridge dynamics, training configuration, and inference settings.

Table 7: Hyperparameter settings of PepHyD.

Name	Value	Description
		Variational Autoencoder
λ_h	0.6	Weight of semantic-latent KL regularization $\mathcal{L}_{\mathrm{KL}}^{h}$.
λ_x	0.8	Weight of coordinate-latent KL regularization $\mathcal{L}_{\mathrm{KL}}^{x}$.
λ_{seq}	1.0	Weight of residue-type reconstruction loss.
λ_{coord}	1.0	Weight of atomic coordinate reconstruction loss.
λ_{local}	0.5	Weight of local-distance reconstruction loss.
Embedding size	512	Dimension of residue and atom embeddings.
Hidden size	512	Hidden dimension of the VAE encoder and decoder.
Semantic latent size	8	Dimension of the residue-level semantic latent $\mathbf{z}_{h,i}$.
Edge size	64	Dimension of edge representations in the equivariant network.
Encoder/decoder layers	6	Number of EPT layers in both the encoder and decoder.
Attention heads	8	Number of attention heads in each EPT layer.
		Hybrid Diffusion–Bridge Dynamics
Hidden size	512	Hidden dimension of the shared denoising network $\mathbf{F}_\theta$.
Denoising layers	6	Number of EPT layers in the shared denoising network.
Attention heads	8	Number of attention heads in each denoising layer.
Total trajectory length T	300	Number of discrete states in the hybrid generative trajectory.
Diffusion ratio ρ_D	0.3	Fraction of the trajectory assigned to diffusion dynamics.
Diffusion steps	90	Number of discrete diffusion steps in the default trajectory.
Bridge steps	210	Number of discrete Schrödinger bridge steps in the default trajectory.
Switch index K	210	Shared boundary between the diffusion and bridge regimes.
Bridge sampling probability p_B	0.7	Probability of sampling a training time from the bridge segment.

Algorithm 1 Training of PepHyD

Require: Peptide graph $\mathcal{G}^p$, target binding-site graph $\mathcal{G}^b$, encoder $\mathcal{E}_\phi$, denoising network $\mathbf{F}_\theta$.

Encode $(\mathcal{G}^p, \mathcal{G}^b)$ with $\mathcal{E}_\phi$ to obtain $\mathbf{U}^p$ and $\mathbf{U}^b$.
Obtain the clean peptide latent $\mathbf{Z}_0 = \mathbf{Z}^p$ through the VAE reparameterization.
repeat
  Select $r \in \{\mathrm{D}, \mathrm{B}\}$ and a corresponding time t.
  Sample $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
  if $r = \mathrm{D}$ **then**
    $a_t^{(r)} \leftarrow a_t^{\mathrm{D}} = \sqrt{\bar{\alpha}_t}$.
    $V_t^{(r)} \leftarrow V_t^{\mathrm{D}} = 1 - \bar{\alpha}_t$.
  else
    $a_t^{(r)} \leftarrow a_t^{\mathrm{B}} = \lambda_t^0 + \lambda_t^1 \sqrt{\bar{\alpha}_K}$.
    $V_t^{(r)} \leftarrow V_t^{\mathrm{B}} = (\lambda_t^1)^2 (1 - \bar{\alpha}_K) + \rho_t^2$.
  end if
  Construct $\mathbf{Z}_t = a_t^{(r)} \mathbf{Z}_0 + \sqrt{V_t^{(r)}}\, \boldsymbol{\epsilon}$.
  Compute η_t^{in}, η_t^{pred}, and η_t^{res} from the statistics of $(\mathbf{Z}_0, \mathbf{Z}_t)$.
  Compute

$$\mathcal{L}_{\mathrm{gen}} = \left\| \mathbf{F}_\theta \left(\eta_t^{\mathrm{in}} \mathbf{Z}_t, t, \mathbf{U}^b \right) - \frac{\eta_t^{\mathrm{pred}} \mathbf{Z}_t - \mathbf{Z}_0}{\eta_t^{\mathrm{res}}} \right\|_2^2 .$$

  Update θ by minimizing $\mathcal{L}_{\mathrm{gen}}$.
until convergence

Algorithm 2 Sampling with PepHyD

Require: Target representation $\mathbf{U}^b$, denoising network $\mathbf{F}_\theta$, decoder $\mathcal{D}_\xi$, T, and K.

Sample $\mathbf{Z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

▷ Diffusion dynamics

for $t = T, \ldots, K+1$ **do**

  Estimate the clean latent:

$$\widehat{\mathbf{Z}}_0 = \eta_t^{\text{pred}} \mathbf{Z}_t - \eta_t^{\text{res}} \mathbf{F}_\theta \left(\eta_t^{\text{in}} \mathbf{Z}_t, t, \mathbf{U}^b \right).$$

  Sample

$$\mathbf{Z}_{t-1} \sim p_{\text{D}} \left(\mathbf{Z}_{t-1} \mid \mathbf{Z}_t, \widehat{\mathbf{Z}}_0 \right).$$

end for

▷ Schrödinger bridge dynamics

Initialize the bridge with the shared boundary $\mathbf{Z}_K$.

for $t = K, \ldots, 1$ **do**

  Estimate the clean latent:

$$\widehat{\mathbf{Z}}_0 = \eta_t^{\text{pred}} \mathbf{Z}_t^{\text{B}} - \eta_t^{\text{res}} \mathbf{F}_\theta \left(\eta_t^{\text{in}} \mathbf{Z}_t^{\text{B}}, t, \mathbf{U}^b \right).$$

  Sample

$$\mathbf{Z}_{t-1}^{\text{B}} \sim p_{\text{B}} \left(\mathbf{Z}_{t-1}^{\text{B}} \mid \mathbf{Z}_t^{\text{B}}, \widehat{\mathbf{Z}}_0 \right).$$

end for

Obtain the final clean latent $\mathbf{Z}_0$.

Decode $\widehat{\mathcal{G}}^p = \mathcal{D}_\xi(\mathbf{Z}_0, \mathbf{U}^b)$.

return $\widehat{\mathcal{G}}^p$.