WHAT LIMITS LATENT DIFFUSION FOR RECEPTOR-CONDITIONED PEPTIDE DESIGN? FROM LATENT SPACE QUALITY TO INTERFACE GUIDANCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Receptor-conditioned peptide design requires sequences and all-atom conformations that fit a target protein interface. Existing methods, however, often rely on noisy latent sampling or uniform diffusion schedules that do not distinguish interface-critical residues from flexible regions. We introduce PepMALD, a manifold-anchored latent diffusion framework for interface-guided peptide design. Its Manifold-Anchored Hierarchical Autoencoder maps designable residues to deterministic latent states that capture local geometry, multi-level interactions, and complex-level context. Latent representation alignment and interface-guided adaptive diffusion then regulate denoising according to receptor proximity. This spatially adaptive process stabilizes interface residues earlier while preserving flexibility away from the binding surface. A dedicated Post-2026 set is designed to evaluate generalization to newly released receptor-peptide complexes. Across multiple benchmarks, PepMALD generates candidates with improved physical plausibility and interface compatibility. Mechanistic analyses associate these gains with a stable latent manifold and interface-first denoising, clarifying how receptor geometry guides residue-level generation. A colorectal cancer case study further indicates its potential for disease-relevant peptide design.

1 INTRODUCTION

Peptides mediate a wide range of protein interactions and constitute an important therapeutic modality between small-molecule drugs and antibodies, owing to their relatively large binding interfaces, tunable sequences, and compatibility with chemical modification (Xiao et al., 2025). Compared with small molecules, peptides are better suited for targeting shallow, extended, or dynamic protein interfaces. Compared with antibodies, they generally have lower structural complexity and are easier to engineer (Callahan et al., 2024).

Early computational peptide design relied on structural search. Candidate conformations were generated through template-based modeling, fragment assembly, or molecular docking, and energy functions were then used to identify plausible binding modes. Although these approaches explicitly model receptor geometry and physical interactions, they require extensive sampling and remain sensitive to initial conformations and scoring functions (Chen et al., 2024). Sequence-based generative methods instead learn amino acid patterns to explore sequence space efficiently, but sequence modeling alone cannot ensure favorable receptor-bound conformations (Li et al., 2025b). Recent structure-based methods incorporate the three-dimensional geometry of receptor-peptide complexes into the generation process and jointly model peptide sequences and conformations using equivariant graph neural networks and diffusion models, which provides a more direct representation of the physical interactions between receptors and peptides (Yim et al., 2023; Lin et al., 2024). This design requires chemically valid sequences and compatible conformations that support stable and specific receptor interactions. Latent diffusion tackles this task in two stages: an autoencoder compresses the receptor-peptide complex into a compact latent representation, and a latent space denoiser recovers peptide sequences and conformations from progressively perturbed states. This formulation reduces the complexity of full-atom generation while retaining joint sequence-structure modeling.

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.

However, generation quality remains limited by weak coordination between latent space construction and reverse denoising. First, variational autoencoders often produce collapsed representations with inactive dimensions and limited effective dimensionality. Their residue-level latents may also fail to retain decodable receptor-interface information, weakening the preservation of peptide structures and interface semantics. Second, the encoding and denoising stages lack an explicit representation constraint. A denoiser trained only for noise prediction can therefore drift away from the encoder-defined latent manifold and lose encoded information during diffusion. Finally, standard diffusion schedules apply uniform noise levels and synchronized updates to all residues, ignoring interface heterogeneity and preventing preferential stabilization of key interfacial residues.

In this paper, we propose PepMALD, an interface-guided peptide design framework based on manifold-anchored latent diffusion. Our main contributions are summarized as follows:

1. A Manifold-Anchored Hierarchical Autoencoder (MHAE) is introduced, in which each peptide residue under design is deterministically mapped to a point on the latent manifold rather than sampled from a posterior distribution. This design reduces sampling noise and the collapse of latent dimensions. Hidden states from multiple encoder layers are further aggregated to preserve local geometry and intermediate interaction patterns, while complex-level global context is incorporated so that interface information is retained during latent space construction.
2. Statistical normalization of the latent variables is combined with intermediate representation alignment through Latent Representation Alignment (LRA). The denoiser's internal residue representations are explicitly anchored to the normalized latent references produced by MHAE, guiding reverse diffusion back toward the latent manifold rather than relying solely on noise prediction.
3. The conventional uniform diffusion schedule is extended by Interface-Guided Adaptive Diffusion (IGAD) to a residue-level adaptive formulation governed by interface distance. The globally uniform Markov chain is thereby extended to a spatially nonuniform process across residues. By adjusting noise intensity and time conditioning in the structural latent space, more unperturbed signal is preserved and the generation trajectories of residues near the receptor interface are stabilized earlier, while greater exploration is retained for residues farther from the interface.
4. The Post-2026 set is constructed to evaluate generalization to newly released complexes, and further validate its applicability through an APC-β-catenin case study.

2 Related Work

Structure-based search methods. These methods formulate peptide design as a conformational search problem on the receptor surface. Rosetta FlexPepDock ab-initio performs flexible peptide docking within the Rosetta energy framework (Raveh et al., 2011). CABS-dock identifies peptide binding modes without prior knowledge of the binding site (Kurcinski et al., 2015), whereas HPEPDOCK improves global-search efficiency through hierarchical docking (Zhou et al., 2018). By incorporating receptor geometry and physics-based energy functions, these methods remain useful for interpreting binding modes and refining predicted structures. However, they require extensive conformational sampling and are sensitive to the initial conformation, conformational-library coverage, and scoring function.

Sequence-based methods. These methods design peptides directly in amino acid sequence space. SaLT&PepPr uses protein language models to predict interaction interfaces and supports the design of peptide-guided protein degraders (Brixi et al., 2023). PepMLM generates linear peptide binders conditioned on a target sequence through masked language modeling (Chen et al., 2025), while EvoBind2 extends sequence-based design to both linear and cyclic peptide binders (Li et al., 2025b). Avoiding explicit structural search enables efficient exploration of sequence space, but sequence modeling alone cannot ensure a favorable receptor-bound conformation. Subsequent structure prediction or docking can screen candidates, but does not impose three-dimensional interface constraints during generation.

Structure-based methods. These methods can be divided into data-space and latent space approaches. Data-space methods jointly model peptide sequences and conformations in three-

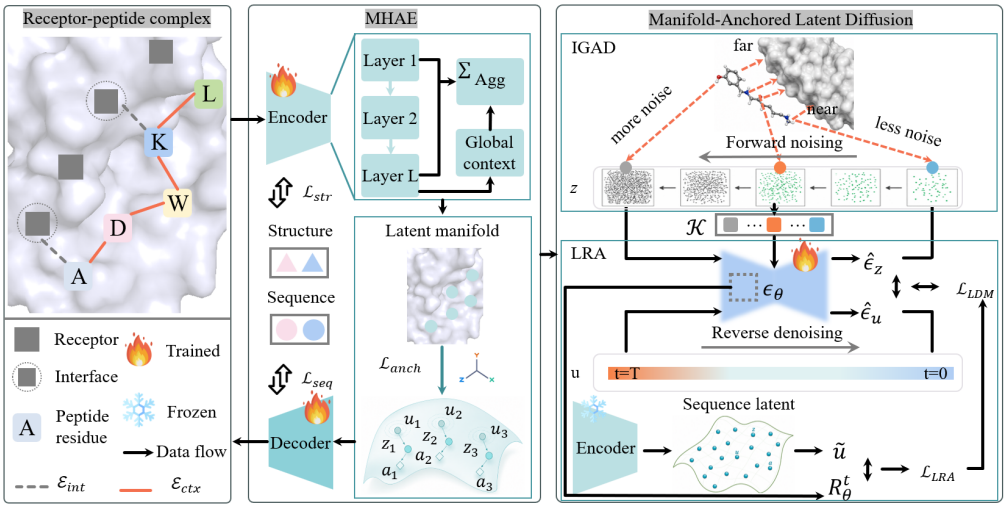

Figure 1: Overall architecture of PepMALD.

dimensional structural space. PepHAR samples interaction hot spots and autoregressively extends peptide fragments under geometric constraints (Li et al., 2025a). DiffPepBuilder uses an SE(3)-equivariant diffusion model to co-generate peptide sequences and structures from target structures and binding-site information (Wang et al., 2024). PPFlow formulates peptide conformation generation as target-aware torsional flow matching on a torus manifold (Lin et al., 2024). PepFlow jointly models backbone rigid frames, side-chain angles, and amino acid types through multimodal flow matching, enabling receptor-conditioned full-atom peptide design (Li et al., 2024). Although these methods directly represent peptide geometry, generation and sampling in high-dimensional full-atom space remain computationally demanding.

Latent space methods first compress receptor-peptide complexes into lower-dimensional representations, a strategy also explored in protein-scale latent diffusion and partially latent flow matching (Li et al., 2026; Geffner et al., 2026). PepGLAD derives residue-level geometric latent variables using a full-atom autoencoder and performs latent diffusion in a receptor-normalized coordinate system (Kong et al., 2024). PepMimic converts interface patterns from known protein binders into constraints for short-peptide design (Kong et al., 2025). PepFGLD incorporates conformational flexibility into diffusion to learn distributions beyond a single static binding conformation (Niu et al.). PepTri introduces physical, evolutionary, and mutual-information signals into denoising to improve peptide stability (Nguyen et al., 2026). These methods improve the efficiency and scalability of full-atom peptide design, but remain dependent on latent space quality and consistency between the encoder-defined representation and the reverse denoising trajectory.

3 Methodology

The architecture of PepMALD is illustrated in Figure 1. Given a receptor-peptide complex, we construct residue-level graphs while retaining full-atom coordinates. MHAE then performs multi-channel equivariant message passing and aggregates hidden states from multiple encoder layers with global context. Each residue under design is encoded as deterministic sequence and structural latent points, avoiding posterior-sampling noise while preserving interface information. A latent diffusion model is trained on these states to jointly generate peptide sequences and structures. LRA anchors intermediate denoiser representations to the normalized sequence latent variables produced by MHAE, while IGAD replaces the globally uniform structural diffusion schedule with an interface-distance-dependent residue-level schedule. Together, these modules constrain reverse diffusion for receptor-conditioned, interface-guided peptide design.

3.1 Manifold-Anchored Hierarchical Autoencoder

Existing peptide models typically construct residue-level latent variables through posterior sampling. The resulting random perturbations are carried into the subsequent latent diffusion process, weakening the preservation of residue-level geometry and receptor-peptide interface context. To ad-

dress this issue, MHAE is introduced to integrate multilevel geometric information during complex-graph encoding and deterministically map each peptide residue under design to a point on the latent manifold. A stable manifold foundation is thereby provided for latent diffusion. The input graph construction and peptide-region masking are detailed in Appendix C.1.

Conventional autoencoders typically construct latent variables using only the final-layer hidden states, which may discard local geometric information captured by earlier layers and interaction patterns learned at intermediate depths. In MHAE, latent representations are instead constructed from residue states across multiple encoder layers. Given the L residue states produced by the encoder, $\{h_i^{(1)}, h_i^{(2)}, \ldots, h_i^{(L)}\}$, hierarchical aggregation is first applied to the hidden states, after which complex-level global context is incorporated:

$$\widehat{h}_i = W_{\text{out}} \left(\frac{1}{L} \sum_{\ell=1}^{L} h_i^{(\ell)} + \frac{1}{|\mathcal{V}|} \sum_{j \in \mathcal{V}} h_j^{(L)} \right) \tag{1}$$

where $h_i^{(\ell)}$ denotes the residue state of residue i at encoder layer ℓ; its construction through multi-channel equivariant message passing is detailed in Appendix C.2.

During latent variable construction, the conventional posterior distribution $q_\phi(z_i \mid \mathcal{G}) = \mathcal{N}(\mu_i, \text{diag}(\sigma_i^2))$ is not used, and no samples are drawn through reparameterization. Instead, the enhanced residue representation is directly projected into a deterministic sequence latent variable, while a residue-level structural latent point is extracted from the encoded structure:

$$u_i = W_u \widehat{h}_i, z_i = \frac{\sum_{c=1}^{C} \widehat{m}_{i,c} X_{i,c}^{\text{enc}}}{\sum_{c=1}^{C} \widehat{m}_{i,c}}, Z_i = (u_i, z_i), \qquad Z_p = \{Z_i \mid i \in \mathcal{V}_p\} \tag{2}$$

where $X_{i,c}^{\text{enc}}$ denotes the coordinate of atom channel c for residue i produced by the encoder, and $\widehat{m}_{i,c}$ denotes the visibility mask of the corresponding atom channel during encoding. The variables u_i and z_i denote the sequence and structural latent points, respectively. Because both u_i and z_i are deterministically produced by the encoder, each peptide residue under design is represented by MHAE as a fixed point on the latent manifold rather than as a variable sampled from a stochastic posterior distribution. The encoder coordinate and visibility-mask construction used in equation 2 are detailed in Appendix C.2.

To obtain a compact and continuous latent space that remains consistent with the underlying residue geometry, a deterministic manifold-anchoring regularizer is introduced in MHAE. The magnitude of the sequence latent point and the displacement of the structural latent point from its geometric anchor are jointly constrained as follows:

$$\mathcal{L}_{\text{anch}} = (1-\eta)\beta_H \left[\frac{1}{2|\mathcal{V}_p|} \sum_{i \in \mathcal{V}_p} \|u_i\|_2^2 \right] + \eta \beta_Z \left[\frac{1}{2|\mathcal{V}_p|} \sum_{i \in \mathcal{V}_p} \|z_i - a_i\|_2^2 \right] \tag{3}$$

where β_H and β_Z control the anchoring strengths of the sequence and structural latent points, respectively, while $a_i = X_{i,C_\alpha}$ denotes the geometric anchor defined by the ground-truth C_α coordinates. The coefficient η controls the balance between sequence and structural reconstruction. The interpretation of the deterministic manifold-anchoring regularizer is provided in Appendix C.3.

During decoding, the input types of the peptide residues are replaced with latent tokens, their coordinates are initialized with z_i, and u_i is injected at the corresponding residue positions. The peptide sequence and full-atom structure are then reconstructed by the decoder under the receptor condition $p_\theta(s_p, X_p \mid Z_p, \mathcal{G}_r) = p_\theta(s_p \mid Z_p, \mathcal{G}_r) p_\theta(X_p \mid s_p, Z_p, \mathcal{G}_r)$, allowing sequence prediction and structural reconstruction to share the same deterministic latent states, so that both residue-type generation and full-atom geometric reconstruction are supported by the latent variables. $\mathcal{G}_r$ denotes the receptor condition defined by the receptor part of the complex graph in Appendix C.1.

The MHAE training objective combines sequence reconstruction, structural reconstruction, and manifold anchoring:

$$\mathcal{L}_{\text{MHAE}} = (1-\eta)\mathcal{L}_{\text{seq}} + \eta \mathcal{L}_{\text{str}} + \mathcal{L}_{\text{anch}} \tag{4}$$

where $\mathcal{L}_{\text{seq}}$, $\mathcal{L}_{\text{str}}$, and $\mathcal{L}_{\text{anch}}$ denote the sequence reconstruction, structural reconstruction, and manifold-anchoring terms, respectively. The explicit definitions of $\mathcal{L}_{\text{seq}}$ and $\mathcal{L}_{\text{str}}$ are provided in Appendix C.4.

3.2 LATENT REPRESENTATION ALIGNMENT

Although each peptide residue under design is mapped by MHAE to a deterministic point on the latent manifold, the reverse trajectory of the subsequent latent diffusion model may still deviate from the latent manifold at intermediate denoising steps when training relies solely on noise prediction. To address this issue, statistical normalization of the latent variables during encoding is combined with intermediate representation alignment during denoising through LRA. Accordingly, the MHAE output u_i is normalized as the sequence-latent reference $\widetilde{u}_i$ for intermediate representation alignment. Its construction and forward perturbation are detailed in Appendix D.1 and Appendix D.2, respectively.

During denoising, let $R_{\theta,i}^{t,\ell_r}$ denote the intermediate representation of residue i produced by denoiser layer ℓ_r at step t. The representation from the last intermediate layer is used for alignment. LRA introduces a learnable projection W_r that maps the intermediate denoiser representation into the normalized sequence-latent space constructed by MHAE:

$$\mathcal{L}_{\mathrm{LRA}} = \frac{1}{|\mathcal{V}_p|} \sum_{i \in \mathcal{V}_p} \left\| W_r \left(R_{\theta,i}^{t,\ell_r} \right) - \widetilde{u}_i \right\|_2^2 \tag{5}$$

The fixed sequence-latent reference $\widetilde{u}_i$ guides the denoiser beyond noise regression without modifying the MHAE encoder. LRA therefore aligns intermediate representations with the sequence-latent space, while the structural latent coordinates are regulated separately by IGAD.

3.3 INTERFACE-GUIDED ADAPTIVE DIFFUSION

Conventional latent diffusion models apply the same noise intensity to all residues under design at a given time step t, so that every residue shares the same cumulative signal-retention coefficient $\overline{\alpha}_t$. To address this limitation, Interface-Guided Adaptive Diffusion (IGAD) is introduced, in which each residue is assigned an interface-conditioned local diffusion coefficient. The diffusion process is thereby extended from a globally uniform Markov chain to a spatially nonuniform residue-level process.

Specifically, an interface anchoring score is estimated in the normalized coordinate system of the complex. Let $\widetilde{z}_i^s$ denote the structural latent coordinate of peptide residue i, where $s = 0$ during training and $s = t$ during reverse sampling. For receptor residue j, q_j denotes its fixed residue-center coordinate. The interface distance and anchoring score are defined as

$$d_i^s = \min_{j \in \mathcal{V}_r} \left\| \widetilde{z}_i^s - q_j \right\|_2, \chi_i^s = \mathbb{I}\left(i \in \mathcal{V}_p\right) \exp\left(-\frac{(d_i^s)^2}{2\sigma_a^2} \right), \qquad \chi_i^s \in [0,1] \tag{6}$$

where σ_a denotes the interface receptive radius, and $\mathbb{I}$ is the indicator function. χ_i^0 during training and χ_i^t during reverse sampling are detailed in Appendix E.1.

To translate the interface anchoring score into an adaptive diffusion schedule, a residue-level schedule compression exponent is defined as follows:

$$\mathcal{K}_i = 1 - \chi_i(1 - \mathcal{K}_{\min}), \qquad \mathcal{K}_i \in [\mathcal{K}_{\min}, 1], \qquad 0 < \mathcal{K}_{\min} \leq 1 \tag{7}$$

Accordingly, the cumulative signal coefficient of standard diffusion is generalized to the residue-adaptive form $\overline{\alpha}_{t,i}^{\mathrm{IGAD}} = (\overline{\alpha}_t)^{\mathcal{K}_i}$. Because $0 < \overline{\alpha}_t < 1$ and $\mathcal{K}_i \leq 1$, we have $\overline{\alpha}_{t,i}^{\mathrm{IGAD}} = (\overline{\alpha}_t)^{\mathcal{K}_i} \geq \overline{\alpha}_t$.

During training, the forward noising process for the structural latent variables under IGAD is defined as follows:

$$\widetilde{z}_i^t = \sqrt{\overline{\alpha}_{t,i}^{\mathrm{IGAD}}}\, \widetilde{z}_i^0 + \sqrt{1 - \overline{\alpha}_{t,i}^{\mathrm{IGAD}}}\, \epsilon_{i,Z}^t, \qquad \epsilon_{i,Z}^t \sim \mathcal{N}\left(0, I_{d_Z}\right) \tag{8}$$

where d_Z denotes the dimension of the structural latent coordinate.

The sequence latent variables continue to follow the standard diffusion schedule in equation 32, while the interface-adaptive schedule is incorporated into the residue-level time conditioning of the denoiser. Under IGAD, the latent diffusion model jointly optimizes noise regression for the sequence and structural latent variables together with LRA:

$$\mathcal{L}_{LDM} = \mathbb{E}_{t,\mathcal{G},\epsilon_H,\epsilon_Z} \left[\omega_H \left\| \epsilon_H^t - \widehat{\epsilon}_H^t \right\|_2^2 + \left\| \epsilon_Z^t - \widehat{\epsilon}_Z^t \right\|_2^2 + \lambda_R \mathcal{L}_{LRA} \right] \tag{9}$$

The first two terms are the sequence and structural noise-regression losses, respectively. The sequence branch follows the standard schedule, whereas the structural branch uses the interface-adaptive perturbation. The full objective and diffusion process are detailed in Appendix E.2-E.5.

4 EXPERIMENTS

4.1 OVERALL PERFORMANCE AND GENERALIZATION

Experiments are conducted on PepBench and PepBDB, with LNR used as the unseen-receptor test set for PepBench. PepMALD is compared with HSRN, dyMEAN, PepGLAD, PepMimic, and PepTri using ΔG, DockQ, RMSD, Success, Div, and Con. Complete experimental settings are provided in Appendix F.

Table 1: Evaluation on sequence-structure co-design. Test ΔG is the mean Rosetta binding energy of native complexes.

	PepBench						PepBDB					
Model	Div ↑	Con ↑	ΔG ↓	Success ↑	DockQ ↑	RMSD ↓	Div ↑	Con ↑	ΔG ↓	Success ↑	DockQ ↑	RMSD ↓
Test	–	–	-37.73	97.84%	0.9999	–	–	–	-36.32	97.36%	0.9999	–
HSRN	0.1580	0.0000	> 0	10.46%	–	–	0.1110	0.0000	> 0	10.86%	–	–
dyMEAN	0.1500	0.0000	-2.26	14.60%	–	–	0.1500	0.0000	-1.92	6.26%	–	–
PepGLAD	0.5677	0.7648	-13.46	47.12%	0.6438	3.6099	0.7985	0.9278	-19.57	47.17%	0.4502	6.1244
PepMimic	0.5527	0.7829	-11.21	52.28%	0.6323	3.5321	0.8039	**0.9573**	-19.27	48.69%	0.4348	6.2147
PepTri	**0.6083**	0.7882	-12.42	48.25%	0.6023	3.8748	**0.8638**	0.9458	-13.90	31.34%	0.3559	9.3298
PepMALD	0.4918	**0.8036**	**-22.72**	**61.10%**	**0.6888**	**2.8421**	0.8074	0.9360	**-19.67**	**49.25%**	**0.4519**	**6.0000**

Sequence-Structure Co-Design. Table 1 reports the results for sequence-structure co-design. In this task, peptide sequences and full-atom conformations are generated jointly under receptor conditioning, which allows receptor modeling and interface geometry to be assessed jointly, unlike sequence generation or structure recovery alone. On PepBench, PepMALD achieves clear improvements over both groups of baselines in binding stability and structural quality, with a higher Success rate and lower structural deviation. These results suggest that plausible sequence-structure combinations can be generated for unseen receptor targets rather than simply reproducing patterns observed in the training data. The differences among methods are smaller on PepBDB, but the overall trend remains similar to that observed on PepBench. PepMALD retains high Success and interface structural quality, indicating that its performance is not limited to a particular test-set configuration and generalizes to a nonredundant split constructed from an independently collected database. The main advantage of PepMALD lies in the validity and stability of its generated candidates. Additional analysis of candidate diversity and usability is provided in Appendix F.4.

Detailed ablation and sensitivity analyses of PepMALD are provided in Appendix K-L. Across secondary-structure, length, sampling-budget, and flexibility groups, PepMALD maintains favorable binding stability and generally outperforms the baselines. Its advantage persists for long and high-flexibility peptides and after controlling for peptide length (Appendix M-N). Chai-1 and Boltz show that PepMALD designs have better backbone and interface-contact recoverability. Rosetta relax and restrained OpenMM simulations further show favorable post-relax energies and stronger preservation of the generated conformations and contacts (Appendix O-P).

Binding Conformation Generation. Figure 2A reports the results for Binding Conformation Generation. Unlike sequence-structure co-design, this task assumes that the peptide sequence is known and requires only its receptor-bound conformation to be generated. It therefore provides a more direct assessment of whether a plausible binding pose can be recovered under a fixed sequence constraint, without relying on sequence optimization to compensate for structural errors. On PepBench, PepMALD achieves the best overall performance. The generated conformations are closer to the reference structures and exhibit favorable binding energies and a higher Success rate. On PepBDB, where cross-receptor generalization is less challenging, PepMALD remains competitive overall and performs particularly well in candidate validity and binding stability. These results suggest that physically plausible binding poses with consistent interface geometry can be generated for unseen receptor targets.

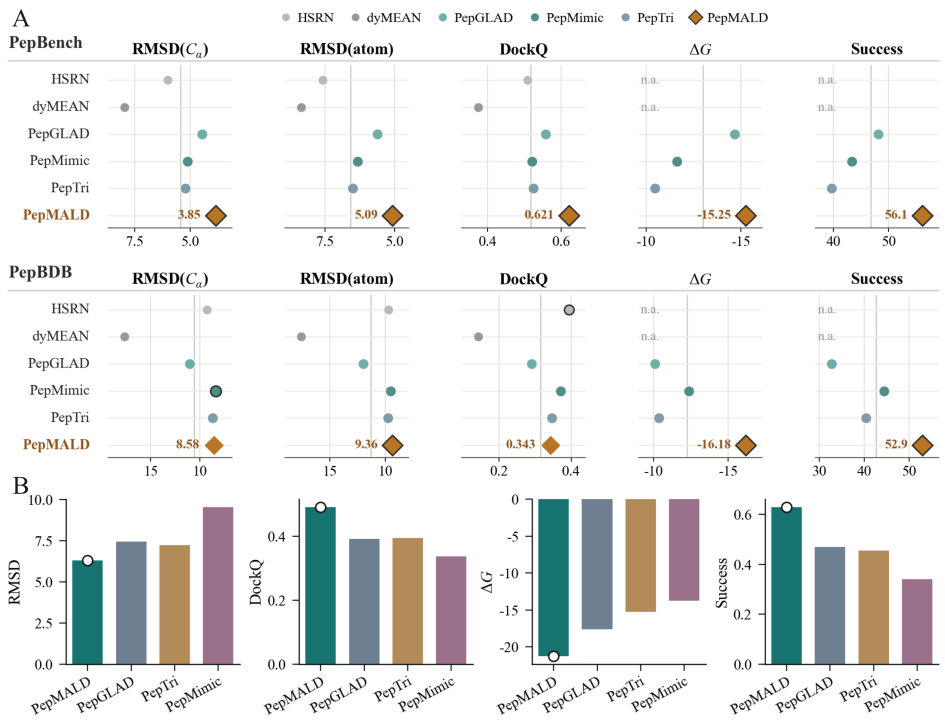

Figure 2: **A.** Evaluation on Binding Conformation Generation. **B.** Model performance on the Post-2026 set.

Generalization to Newly Released Complexes. To assess generalization beyond established benchmark splits, we construct the Post-2026 test set, its construction and characteristics are detailed in Appendix F.6. The same evaluation protocol is applied to PepMALD and all baselines on the Post-2026 set. As shown in Figure 2B, 40 candidate peptides are generated for each target, and PepMALD achieves the best overall performance across the evaluated metrics. Its median RMSD is 6.30 Å, compared with 7.23 Å for PepTri. PepMALD also achieves the highest mean DockQ and Success and the most favorable mean ΔG. These results indicate that its performance advantage extends beyond the original benchmark splits to post-cutoff complexes with different peptide-length and sequence-composition distributions.

4.2 Analysis of Manifold-Anchored Latent Space Quality

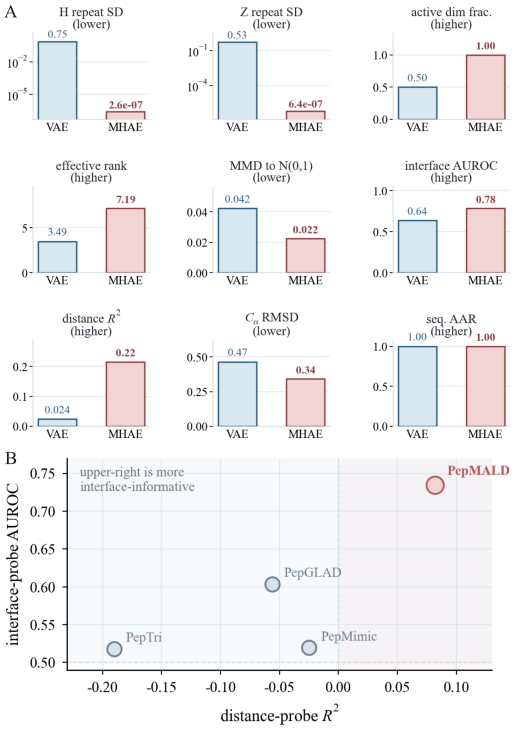

Figure 3: **A.** Direct VAE-to-MHAE comparison on latent space analysis. **B.** Linear-probe analysis of interface information in frozen residue latents.

The generative performance of a latent diffusion model depends strongly on the latent space constructed by its autoencoder. To examine the latent space constructed by MHAE, we compare it with a conventional VAE in terms of encoding stability, dimensional utilization, prior compatibility, interface information, and reconstruction quality.

As shown in Figure 3A, the standard deviations obtained from repeated sampling of H and Z are 0.75 and 0.53 for the conventional VAE, respectively, but decrease to approximately 2.6×10^{-7} and 6.4×10^{-7} with MHAE. The conventional VAE has an active dimension fraction of 0.5 and an effective rank of 3.49, whereas all eight dimensions are active in MHAE and the effective rank increases to 7.19. MHAE also reduces the MMD to the standard normal distribution from 0.042 to 0.022. These improvements are achieved without sacrificing autoencoder reconstruction, as MHAE reduces the deterministic decoding C_α RMSD from 0.47 to 0.34 while maintaining similar sequence reconstruction accuracy.

More receptor-peptide interface information is also retained in the latent representations produced by MHAE. Frozen-latent linear probing shows that MHAE increases the interface contact AUROC from 0.64 to 0.78 and the receptor-distance R^2 from 0.02 to 0.22. In the cross-model comparison shown in Figure 3B, PepMALD achieves an interface AUROC of 0.74, higher than those of the VAE-based baselines, and its receptor-distance R^2 is the only clearly positive result. MHAE therefore improves encoding stability, dimensional utilization, prior compatibility, and interface information content while preserving reliable sequence and structure reconstruction. Additional analyses of encoding variability, latent-manifold structure, dimension-wise variance, and prior compatibility are provided in Appendix G.

4.3 Analysis of Latent Diffusion Dynamics

We further examine whether PepMALD changes the reverse dynamics of latent diffusion. Rather than focusing on generation metrics, this analysis addresses two questions. First, does LRA guide the intermediate denoiser representations toward the reference latent state produced by MHAE? Second, does IGAD modulate residue-level local transitions according to receptor-peptide interface anchoring?

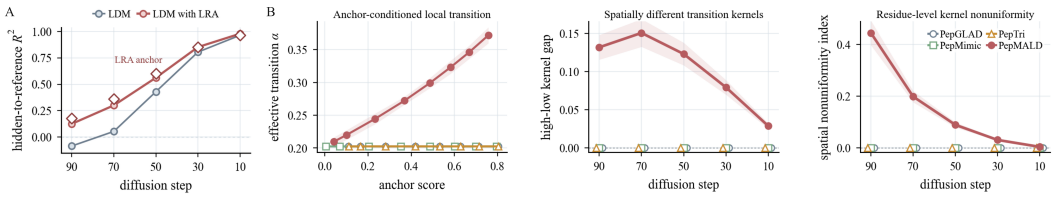

Figure 4: **A.** Hidden-to-reference reconstruction R^2 for the traditional and alignment-enhanced LDMs. **B.** Anchor-dependent local transitions and residue-level Markov heterogeneity under IGAD.

We first analyze the effect of LRA on the intermediate representations of the denoiser. At different reverse diffusion steps, the linear recoverability of the reference latent state from the denoiser hidden state is evaluated using the hidden-to-reference reconstruction R^2. As shown in Figure 4A, LRA improves the hidden-to-reference reconstruction R^2 throughout reverse diffusion compared with the traditional LDM. This result indicates that the latent reference guides the intermediate hidden representations toward the MHAE latent space. The step-wise effect of LRA is further analyzed in Appendix H.

We next examine the spatial nonuniformity introduced by IGAD at the level of the Markov transition kernel. As shown in Figure 4B, at $t = 70$, the effective transition α of PepMALD increases monotonically with the anchor score, whereas those of PepGLAD, PepMimic, and PepTri remain nearly constant. PepMALD also exhibits nonzero high-low transition-kernel gaps and higher residue-level kernel nonuniformity indices across multiple diffusion steps. Both measures remain close to zero for the three baselines. The high-low transition-kernel gap and nonuniformity index gradually decrease as reverse diffusion proceeds, indicating that the spatial differences introduced by IGAD act most strongly during the earlier stages of generation.

Together, these results show that LRA improves the recovery of the reference latent state from the denoiser hidden representation, while IGAD modulates local transition kernels according to interface anchoring. Additional analyses of the intermediate denoising dynamics and residue-level trajectories are provided in Appendix H and Appendix I.

4.4 Case Study: Peptide Generation for Colorectal-Cancer-Relevant APC-β-Catenin Interfaces

PepMALD was further evaluated in a disease-relevant peptide-generation setting centered on the APC-β-catenin axis in colorectal cancer. Two structurally characterized interfaces were selected: the compact APC 15-mer interface from PDB 1JPP and the extended APC-R3 20-mer interface from PDB 1T08. These structures allow peptide generation to be examined on two APC-related β-catenin binding grooves with different lengths and interaction patterns. Candidate-level energy evaluation is provided in Appendix J.

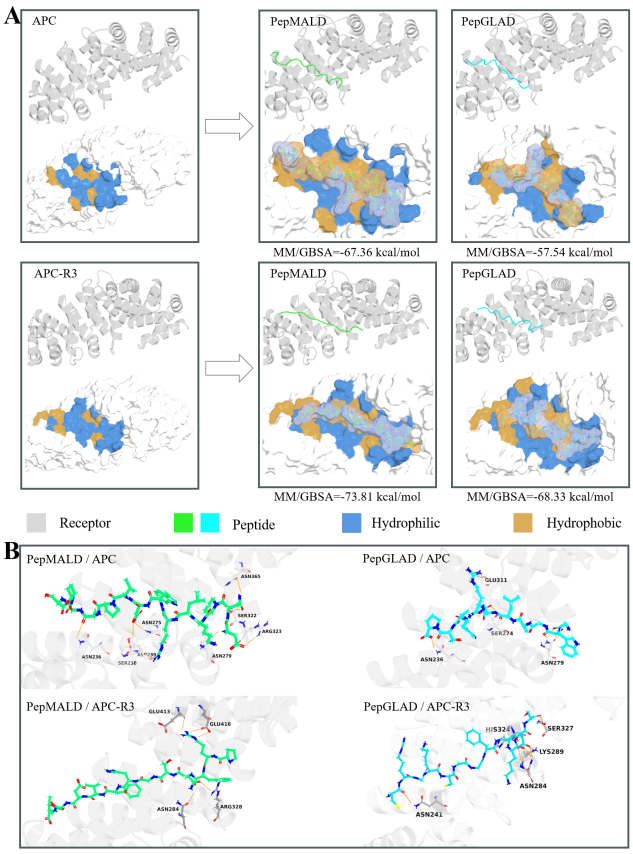

Figure 5: APC-β-catenin case study. **A.** Representative PepMALD (green) and PepGLAD (cyan) binding poses on 1JPP and 1T08. **B.** Corresponding local interactions; orange dashed lines indicate hydrogen bonds or polar contacts.

Representative-pose analysis revealed differences in predicted binding and groove engagement between PepMALD and PepGLAD. As shown in Figure 5A, at the 1JPP interface, the PepMALD peptide achieved an MM/GBSA binding energy of -67.36 kcal/mol, which was 9.82 kcal/mol lower than that of the corresponding PepGLAD peptide. The PepMALD pose followed the local groove geometry and contacted both hydrophilic and hydrophobic surface regions, whereas the PepGLAD pose occupied a smaller region. At the extended 1T08 interface, PepMALD and PepGLAD achieved binding energies of -73.81 and -68.33 kcal/mol, respectively. The PepMALD peptide extended across a longer region of the groove, while the PepGLAD pose remained more localized.

The local interaction patterns showed a similar distinction. As shown in Figure 5B, on 1JPP, the PepMALD peptide formed contacts distributed along the peptide-binding path, providing multiple receptor-facing anchors across the groove. Its contacts on 1T08 were likewise distributed across different regions of the extended interface. By contrast, the PepGLAD interactions were concentrated within more localized regions on both targets. The difference between the models therefore reflected the spatial organization of the contact network rather than contact number alone.

5 Conclusion

This study presents PepMALD, a manifold-anchored latent diffusion framework for receptor-conditioned peptide design. PepMALD replaces stochastic posterior sampling with deterministic latent states and incorporates receptor-peptide geometry through interface-guided adaptive diffusion, stabilizing interface-proximal residues earlier while preserving conformational flexibility. Experiments show improved generation quality, while latent space and diffusion analyses reveal a stable, interface-informative manifold and receptor-proximity-dependent denoising. These results clarify how receptor geometry guides peptide generation rather than acting only as an external condition. The colorectal cancer case study further indicates the potential of PepMALD for disease-relevant peptide candidate discovery.

AI USE STATEMENT

Large Language Models (LLMs) were used to aid in the writing and polishing of the manuscript. Specifically, we used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text. The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. In this study, no human subjects or animal experimentation was involved. All datasets used were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

REPRODUCIBILITY STATEMENT

We have made every effort to ensure that the results presented in this paper are reproducible. Additionally, all datasets used in this paper are publicly available resources, ensuring the consistency and reproducibility of the evaluation results. We believe these measures will enable other researchers to reproduce our work and further advance the field.

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, et al. The rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation*, 13(6):3031, 2017.

Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1): 235–242, 2000.

Garyk Brixi, Tianzheng Ye, Lauren Hong, Tian Wang, Connor Monticello, Natalia Lopez-Barbosa, Sophia Vincoff, Vivian Yudistyra, Lin Zhao, Elena Haarer, et al. Salt&peppr is an interface-predicting language model for designing peptide-guided protein degraders. *Communications Biology*, 6(1):1081, 2023.

Alex J Callahan, Satish Gandhesiri, Tara L Travaline, Rahi M Reja, Lia Lozano Salazar, Stephanie Hanna, Yen-Chun Lee, Kunhua Li, Olena S Tokareva, Jean-Marie Swiecicki, et al. Mirror-image ligand discovery enabled by single-shot fast-flow synthesis of d-proteins. *Nature communications*, 15(1):1813, 2024.

Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, Moises Lopez-Gonzalez, Alexander Mestre, Rio Watson, Sophia Vincoff, Lin Zhao, et al. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, pp. 1–9, 2025.

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.

Tomas Geffner, Kieran Didi, Zhonglin Cao, Danny Reidenbach, Zuobai Zhang, Christian Dallago, Emine Kucukbenli, Karsten Kreis, and Arash Vahdat. La-proteina: Atomistic protein generation via partially latent flow matching. In *International Conference on Learning Representations*, volume 2026, pp. 67350–67395, 2026.

Nam-Chul Ha, Takashi Tonozuka, Jennifer L Stamos, Hee-Jung Choi, and William I Weis. Mechanism of phosphorylation-dependent binding of apc to β-catenin and its role in β-catenin degradation. *Molecular cell*, 15(4):511–521, 2004.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Antibody-antigen docking and design via hierarchical structure refinement. In *International Conference on Machine Learning*, pp. 10217–10227. PMLR, 2022.

Xiangzhe Kong, Wenbing Huang, and Yang Liu. End-to-end full-atom antibody design. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 17409–17429. PMLR, 23–29 Jul 2023.

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, Rui Jiao, Haowei Lin, Ruihan Guo, Wenbing Huang, Wei-Ying Ma, Zihua Wang, Yang Liu, and Jianzhu Ma. Peptide design through binding interface mimicry with pepmimic. *Nature biomedical engineering*, pp. 1–16, 2025.

Mateusz Kurcinski, Michal Jamroz, Maciej Blaszczyk, Andrzej Kolinski, and Sebastian Kmiecik. Cabs-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic acids research*, 43(W1):W419–W424, 2015.

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. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27615–27640. PMLR, 21–27 Jul 2024.

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. In *International Conference on Learning Representations*, volume 2025, pp. 9282–9314, 2025a.

Qiuzhen Li, Efstathios Nikolaos Vlachos, and Patrick Bryant. Design of linear and cyclic peptide binders from protein sequence information. *Communications Chemistry*, 8(1):211, 2025b.

Shaoning Li, Le Zhuo, Yusong Wang, Mingyu Li, Xinheng He, Fandi Wu, Hongsheng Li, and Pheng-Ann Heng. Proteinae: Protein diffusion autoencoders for structure encoding. In *International Conference on Learning Representations*, volume 2026, pp. 69939–69960, 2026.

Haitao Lin, Odin Zhang, Huifeng Zhao, Dejun Jiang, Lirong Wu, Zicheng Liu, Yufei Huang, and Stan Z. Li. PPFLOW: Target-aware peptide design with torsional flow matching. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 30510–30528. PMLR, 21–27 Jul 2024.

Ngoc-Quang Nguyen, Jaeyoon Jung, Seijung Kim, Sunkyu Kim, and Jaewoo Kang. Peptri: Trimodal guided all-atom diffusion for peptide design via physics, evolution, and mutual information. In *The Fourteenth International Conference on Learning Representations*, 2026.

Dongjiang Niu, Xiaofeng Wang, Zhiqiang Wei, and Zhen Li. Flexibility-aware geometric latent diffusion for full-atom peptide design. In *Forty-third International Conference on Machine Learning*.

Barak Raveh, Nir London, Lior Zimmerman, and Ora Schueler-Furman. Rosetta flexpepdock ab-initio: simultaneous folding, docking and refinement of peptides onto their receptors. *PloS one*, 6 (4):e18934, 2011.

Katharine Eklof Spink, Sofiya G Fridman, and William I Weis. Molecular mechanisms of β-catenin recognition by adenomatous polyposis coli revealed by the structure of an apc–β-catenin complex. *The EMBO journal*, 20(22):6203–6212, 2001.

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.

Chai Discovery team, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhonikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life. *BioRxiv*. doi: 10.1101/2024.10.10.615955.

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.

Fanhao Wang, Yuzhe Wang, Laiyi Feng, Changsheng Zhang, and Luhua Lai. Target-specific de novo peptide binder design with diffpepbuilder. *Journal of Chemical Information and Modeling*, 64(24):9135–9149, 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.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, et al. Boltz-1 democratizing biomolecular interaction modeling. *BioRxiv*, 2025. doi: 10.1101/2024.11.19.624167.

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.

Jianyi Yang and Yang Zhang. Protein structure and function prediction using i-tasser. *Current protocols in bioinformatics*, 52(1):5.8.1–5.8.15, 2015.

Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 40001–40039. PMLR, 23–29 Jul 2023.

Pei Zhou, Bowen Jin, Hao Li, and Sheng-You Huang. Hpepdock: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic acids research*, 46(W1):W443–W450, 2018.

A ALGORITHM OF PEPMALD

Algorithm 1 PepMALD: Receptor-Conditioned Peptide Design

Require: Training receptor-peptide complexes $\mathcal{D}_{\text{train}}$; target receptor graph $\mathcal{G}_r$; designed peptide residue set $\mathcal{V}_p$; diffusion schedule $\{\beta_t\}_{t=1}^{T}$; IGAD update interval K

Ensure: Generated peptide sequence $\widehat{s}_p$ and full-atom structure $\widehat{X}_p$

Stage 1: Pre-training the MHAE
for all receptor-peptide complexes $\mathcal{G} \in \mathcal{D}_{\text{train}}$ **do**
  Step 1: Hierarchical Complex Encoding
  Construct the residue-level complex graph and peptide masks { Graph construction }
  Encode residue and atom-channel features through multichannel equivariant message passing
  Aggregate multilevel residue states and complex-level global context { equation 1 }
  Step 2: Deterministic Latent Construction
  Construct the sequence latent point u_i and structural latent point z_i for each $i \in \mathcal{V}_p$ { equation 2 }
  Set $a_i \leftarrow X_{i,C_\alpha}$ as the MHAE geometric anchor
  Step 3: Reconstruction and Manifold Anchoring
  Decode the peptide sequence and full-atom structure from $Z_p = \{(u_i, z_i) \mid i \in \mathcal{V}_p\}$
  Update MHAE using $\mathcal{L}_{\text{MHAE}}$ { equation 4 }
end for
Compute the sequence-latent statistics μ_H and D_H
Freeze the trained MHAE encoder and decoder
Stage 2: Training the Interface-Guided Latent Diffusion Model
for all receptor-peptide complexes $\mathcal{G} \in \mathcal{D}_{\text{train}}$ **do**
  Encode the complex using the frozen MHAE
  Normalize the sequence latent points: $\widetilde{u}_i \leftarrow D_H^{-1}(u_i - \mu_H)$
  Sample a diffusion step t and Gaussian noises $\epsilon_H^t, \epsilon_Z^t$
  Step 1: Interface-Adaptive Forward Perturbation
  Compute the clean-state interface score χ_i^0 from $\widetilde{z}_i^0$ and the receptor coordinates { equation 6 }
  Convert χ_i^0 into the residue-level exponent $\mathcal{K}_i$ { equation 7 }
  Perturb $\widetilde{u}_i$ with the standard schedule and $\widetilde{z}_i^0$ with the IGAD schedule { equation 32 and equation 8 }
  Step 2: Joint Denoising and Representation Alignment
  Predict $\widehat{\epsilon}_H^t, \widehat{\epsilon}_Z^t$, and the intermediate representation R_θ^t
  Align R_θ^t with the fixed sequence-latent reference $\widetilde{u}$ through LRA { equation 5 }
  Update the denoiser using $\mathcal{L}_{\text{LDM}}$ { equation 9 }
end for
Stage 3: Interface-Guided Reverse Sampling
Initialize $\widetilde{u}^T \sim \mathcal{N}(0, I_{d_H})$ and $\widetilde{z}^T \sim \mathcal{N}(0, I_{d_Z})$
for $t = T, T-1, \ldots, 1$ **do**
  Recompute χ_i^t and $\mathcal{K}_i^t$ every K steps { State-adaptive IGAD }
  Predict $\widehat{\epsilon}_{i,H}^t$ and $\widehat{\epsilon}_{i,Z}^t$ with residue-level time conditioning
  Update $\widetilde{u}_i^{t-1}$ using the standard reverse process { equation 48 }
  Update $\widetilde{z}_i^{t-1}$ using the IGAD reverse process { equation 47 }
end for
Map $\widetilde{u}^0$ back to the MHAE latent space and restore the structural coordinates
Decode the final latent states into $(\widehat{s}_p, \widehat{X}_p)$
return $\widehat{s}_p, \widehat{X}_p$

B DEFAULT NOTATION

Calligraphic uppercase letters denote sets, graphs, and objective functions, uppercase Roman letters denote coordinate arrays or linear maps, and lowercase letters denote residue-level variables. Unless otherwise stated, i and j index residues, c and c' index atom channels, ℓ indexes network layers, and t indexes diffusion steps. The notation specific to PepMALD is summarized below.

Indices and Basic Operators

i, j	Residue indices.
$c, c' \in \{1, \ldots, C\}$	Atom-channel indices, where C is the number of atom channels retained for each residue.
$\ell \in \{1, \ldots, L\}$	Encoder-layer index, where L is the number of encoder layers.
$t \in \{1, \ldots, T\}$	Global diffusion step, where T is the total number of diffusion steps.
$\lvert\mathcal{S}\rvert$	Cardinality of a set $\mathcal{S}$.
$\lVert x \rVert_2$	Euclidean norm of x.
$\mathrm{diag}(x)$	Diagonal matrix whose diagonal entries are given by x.
$x \odot y$	Element-wise multiplication of x and y.
$\mathbb{I}[\cdot]$	Indicator function.
$\mathcal{N}(0, I_d)$	The d-dimensional standard Gaussian distribution.
	Complex Graph and Residue Representation
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$	Residue-level graph of a receptor-peptide complex.
$\mathcal{V} = \mathcal{V}_r \cup \mathcal{V}_p$	Node set consisting of receptor residues $\mathcal{V}_r$ and peptide residues under design $\mathcal{V}_p$.
$\mathcal{E} = \mathcal{E}_{\mathrm{ctx}} \cup \mathcal{E}_{\mathrm{int}}$	Edge set consisting of intra-region context edges $\mathcal{E}_{\mathrm{ctx}}$ and receptor-peptide interface edges $\mathcal{E}_{\mathrm{int}}$.
$s_i,\ p_i$	Amino acid type and positional index of residue i, respectively.
$X_i \in \mathbb{R}^{C\times 3}$	Full-atom coordinate array of residue i.
$m_i \in \{0,1\}^C$	Atom mask indicating the valid atom channels of residue i.
$\widetilde{s}_i,\ \rho$	Masked residue type and peptide-region masking ratio, respectively.
$h_i^{(\ell)}$	Hidden state of residue i at encoder layer ℓ.
$X_{i,c}^{(\ell)}$	Coordinate of atom channel c for residue i at encoder layer ℓ.
$r_{ij}^{(\ell)},\ e_{ij}^{(\ell)}$	Geometric relation and edge message between residues i and j at layer ℓ, respectively.
$\phi_e^{(\ell)},\ \phi_h^{(\ell)},\ \psi_x^{(\ell)}$	Geometric message, residue-state update, and coordinate-weight projection functions at layer ℓ, respectively.
$P_c,\ P_X,\ \mathrm{Agg}$	Channel-adaptive pooling, valid-coordinate aggregation, and neighborhood aggregation operators, respectively.
	MHAE Latent Representation
$\widehat{h}_i$	Enhanced residue representation obtained by aggregating encoder states across layers and incorporating complex-level global context.
$W_{\mathrm{out}},\ W_u$	Learnable projections used for hierarchical aggregation and sequence-latent construction, respectively.
$X_{i,c}^{\mathrm{enc}}$	Coordinate of atom channel c for residue i produced by the MHAE encoder.
$\widehat{m}_{i,c}$	Visibility mask of atom channel c during MHAE encoding.
$u_i \in \mathbb{R}^{d_H}$	Deterministic sequence latent point of peptide residue i.
$z_i \in \mathbb{R}^{d_Z}$	Deterministic structural latent point of peptide residue i.
$Z_i = (u_i, z_i), Z_p = \{Z_i \mid i \in \mathcal{V}_p\}$	Joint sequence-structure latent state of residue i and the collection of latent states for the peptide region under design.
$a_i = X_{i,C_\alpha}$	Geometric anchor of residue i, defined by its ground-truth C_α coordinate.
$\mathcal{B}_{bb},\ \mathcal{B}_{sc}$	Sets of backbone and side-chain bonds in the peptide region, respectively.
$\mathcal{T}_{\mathcal{Q}},\ \Omega_{abcd}$	Set of atom quadruplets used for side-chain torsions and the dihedral angle defined by atoms (a, b, c, d), respectively.
	LRA and Standard Diffusion

$\mu_H,\ \sigma_H,\ \overline{\sigma}_H$	Dimension-wise mean, standard deviation, and lower-bounded standard deviation of the sequence latent variables in the training set.
$D_H = \mathrm{diag}(\overline{\sigma}_H)$	Diagonal scaling matrix used to normalize the sequence latent variables.
$\widetilde{u}_i = D_H^{-1}(u_i - \mu_H)$	Normalized sequence-latent reference of residue i.
$\beta_t, \alpha_t = 1 - \beta_t, \overline{\alpha}_t = \prod_{\tau=1}^{t} \alpha_\tau$	Noise intensity, signal-retention coefficient, and cumulative signal-retention coefficient of the standard diffusion schedule.
$\widetilde{u}_i^t,\ \widetilde{z}_i^t$	Sequence and structural latent states of residue i at diffusion step t.
$R_{\theta,i}^{t,\ell_r}$	Intermediate representation of residue i produced by denoiser layer ℓ_r at diffusion step t.
W_r	Learnable projection that maps the intermediate denoiser representation into the normalized sequence-latent space.
$\epsilon_{i,H}^t,\ \epsilon_{i,Z}^t$	Gaussian noise applied to the sequence and structural latent variables, respectively.
$\widehat{\epsilon}_{i,H}^t,\ \widehat{\epsilon}_{i,Z}^t$	Sequence and structural noise predicted by the denoising network, respectively.
	Interface-Guided Adaptive Diffusion
$\widetilde{z}_i^s$	Structural latent coordinate used to estimate interface proximity, where $s = 0$ during training and $s = t$ during reverse sampling.
q_j	Fixed residue-center coordinate of receptor residue j in the normalized coordinate system.
$d_i^s = \min_{j \in \mathcal{V}_r} \lVert \widetilde{z}_i^s - q_j \rVert_2$	Distance from peptide residue i to the closest receptor residue.
χ_i^s	Interface anchoring score of peptide residue i, computed from d_i^s.
σ_a	Interface receptive radius used in the anchoring-score function.
$\mathcal{K}_i = 1 - \chi_i(1 - \mathcal{K}_{\min})$	Residue-level schedule compression exponent, where $\mathcal{K}_i \in [\mathcal{K}_{\min}, 1]$.
$\overline{\alpha}_{t,i}^{\mathrm{IGAD}} = (\overline{\alpha}_t)^{\mathcal{K}_i}$	Residue-adaptive cumulative signal-retention coefficient used by IGAD.
$\beta_{t,i}^{\mathrm{IGAD}} = 1 - (1 - \beta_t)^{\mathcal{K}_i}$	Residue-level noise intensity induced by the interface-adaptive schedule.
$\gamma_i^{\mathrm{IGAD}}(t)$	Residue-level time embedding constructed from $\beta_{t,i}^{\mathrm{IGAD}}$.
ϵ_θ	Parameterized denoising network for predicting sequence and structural noise.
	Objectives and Evaluation Quantities
$\mathcal{L}_{\mathrm{seq}},\ \mathcal{L}_{\mathrm{str}},\ \mathcal{L}_{\mathrm{anch}}$	Sequence-reconstruction, structure-reconstruction, and manifold-anchoring losses of MHAE, respectively.
$\mathcal{L}_{\mathrm{MHAE}}$	Overall MHAE objective combining sequence reconstruction, structural reconstruction, and manifold anchoring.
$\mathcal{L}_{\mathrm{LRA}}$	Representation-alignment loss between the projected denoiser representation and the normalized sequence-latent reference.
$\mathcal{L}_{\mathrm{LDM}}$	Latent-diffusion objective combining sequence noise regression, structural noise regression under IGAD, and LRA.
$\eta,\ \beta_H,\ \beta_Z$	Balance between sequence and structural reconstruction and the anchoring strengths of the sequence and structural latent points.
$\lambda_X, \lambda_{C_\alpha}, \lambda_{bb}, \lambda_{sc}, \lambda_{\mathcal{Q}}$	Weights of the full-coordinate, C_α, backbone-bond, side-chain-bond, and torsional reconstruction terms.
$\omega_H,\ \lambda_R$	Weight of the sequence noise-regression term and weight of the LRA term in $\mathcal{L}_{\mathrm{LDM}}$, respectively.
ΔG, DockQ, RMSD	Binding-energy, interface-quality, and structural-deviation measures used in the evaluation.
Success, Div, Con	Overall candidate success, generated-sample diversity, and sequence-structure consistency measures, respectively.

C Details of the Manifold-Anchored Hierarchical Autoencoder

C.1 Receptor-Peptide Complex Graph and Peptide-Region Masking

Given a receptor-peptide complex graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the node set $\mathcal{V} = \mathcal{V}_r \cup \mathcal{V}_p$ consists of receptor residues $\mathcal{V}_r$ and peptide residues under design $\mathcal{V}_p$. The edge set is partitioned as $\mathcal{E} = \mathcal{E}_{\mathrm{ctx}} \cup \mathcal{E}_{\mathrm{int}}$, where intra-region edge set $\mathcal{E}_{\mathrm{ctx}} = \{(i,j) \in \mathcal{E} \mid \mathbb{I}[i \in \mathcal{V}_p] = \mathbb{I}[j \in \mathcal{V}_p]\}$ contains receptor-receptor and peptide-peptide edges, and inter-region edge set $\mathcal{E}_{\mathrm{int}} = \{(i,j) \in \mathcal{E} \mid \mathbb{I}[i \in \mathcal{V}_p] \neq \mathbb{I}[j \in \mathcal{V}_p]\}$

contains receptor-peptide edges. Each residue i is represented by its amino acid type s_i, residue position p_i, full-atom coordinates $X_i \in \mathbb{R}^{C\times 3}$, and atom mask $m_i \in \{0,1\}^C$, where C denotes the number of atom channels. To encourage residue representations to be learned from the complex context and local geometry, the amino acid types in the peptide region are randomly masked, yielding the perturbed input sequence $\widetilde{s}$:

$$\widetilde{s}_i = \begin{cases} [\text{MASK}], & i \in \mathcal{V}_p,\ \xi_i < \rho, \\ s_i, & \text{otherwise}, \end{cases} \qquad \xi_i \sim \mathcal{U}(0,1) \tag{10}$$

where ρ denotes the masking ratio, and ξ_i is sampled independently for each residue i from a uniform distribution $\mathcal{U}$. The initial representation of residue i is constructed from its perturbed residue type and positional index as $h_i^{(0)} = \mathrm{Emb}_s(\widetilde{s}_i) + \mathrm{Emb}_p(p_i)$, where Emb_s and Emb_p denote learnable embeddings.

C.2 Multichannel Equivariant Message Passing

The complex graph is then encoded by MHAE through multichannel equivariant message passing. At the ℓ-th layer, the geometric relation between residues i and j is represented by their coordinate differences across atom channels, from which the edge message is computed as follows:

$$r_{ij}^{(\ell)} = \left[m_{i,c} m_{j,c'} \left\| X_{i,c}^{(\ell)} - X_{j,c'}^{(\ell)} \right\|_2^2 \right]_{c,c'=1}^{C}, e_{ij}^{(\ell)} \quad = \phi_e^{(\ell)} \left(h_i^{(\ell)}, h_j^{(\ell)}, r_{ij}^{(\ell)} \right) \tag{11}$$

The residue hidden states and full-atom coordinates are then updated as follows:

$$\begin{aligned} h_i^{(\ell+1)} &= h_i^{(\ell)} + \phi_h^{(\ell)} \left(h_i^{(\ell)}, \sum_{j:(i,j)\in\mathcal{E}} e_{ij}^{(\ell)} \right), \\ X_{i,c}^{(\ell+1)} &= X_{i,c}^{(\ell)} + \mathrm{Agg}_{j:(i,j)\in\mathcal{E}} \left\langle \left[P_c \left(\psi_x^{(\ell)}(e_{ij}^{(\ell)}); m_i \right) \right]_c \left(X_{i,c}^{(\ell)} - P_X(X_j^{(\ell)}, m_j) \right) \right\rangle \end{aligned} \tag{12}$$

where $\phi_e^{(\ell)}$ denotes the geometric message function, $\phi_h^{(\ell)}$ the residue-state update function, P_c denotes channel-adaptive pooling, and $P_X(X_j^{(\ell)}, m_j)$ aggregates only the valid atom coordinates of residue j. The function $\psi_x^{(\ell)}$ projects edge messages onto coordinate-update weights, while Agg denotes neighborhood aggregation. Unlike methods in which latent variables are constructed from isolated peptide sequences, multichannel equivariant message passing is performed by MHAE over the complete receptor-peptide complex graph, allowing complex-level contextual information to be captured by the peptide latent variables from the outset.

C.3 Interpretation of the Deterministic Manifold-Anchoring Regularizer

The following interpretation applies to the deterministic manifold-anchoring regularizer in equation 3. Unlike KL divergence, which regularizes latent variables toward an isotropic Gaussian prior, this constraint is applied directly to deterministic latent points. The scale of the sequence latent variables is bounded to maintain a compact latent space suitable for diffusion. The structural latent variables are also anchored near the ground-truth residue backbone, thereby reducing geometric drift in the latent space.

Proposition 1. Control of latent compactness and structural drift Let $n = |\mathcal{V}_p|$, assume $0 < \eta < 1$, $\beta_H > 0$, and $\beta_Z > 0$, and suppose that the manifold-anchoring loss in equation 3 satisfies $\mathcal{L}_{\text{anch}} \le \varepsilon_{\text{A}}$. Then the sequence and structural latent points satisfy

$$\frac{1}{n} \sum_{i\in\mathcal{V}_p} \|u_i\|_2^2 \le \frac{2\varepsilon_{\text{A}}}{(1-\eta)\beta_H} \tag{13}$$

and

$$\frac{1}{n} \sum_{i\in\mathcal{V}_p} \|z_i - a_i\|_2^2 \le \frac{2\varepsilon_{\text{A}}}{\eta\beta_Z} \tag{14}$$

Moreover, for any displacement threshold $r > 0$, the fraction of peptide residues whose structural latent point is displaced from its geometric anchor by at least r is bounded by

$$\frac{|\{i \in \mathcal{V}_p : \|z_i - a_i\|_2 \geq r\}|}{|\mathcal{V}_p|} \leq \min\left\{1, \frac{2\varepsilon_{\mathrm{A}}}{\eta\beta_Z r^2}\right\} \tag{15}$$

Proof. Both terms in $\mathcal{L}_{\mathrm{anch}}$ are nonnegative. Therefore,

$$\mathcal{L}_{\mathrm{anch}} \geq \frac{(1-\eta)\beta_H}{2n} \sum_{i \in \mathcal{V}_p} \|u_i\|_2^2 \tag{16}$$

and

$$\mathcal{L}_{\mathrm{anch}} \geq \frac{\eta\beta_Z}{2n} \sum_{i \in \mathcal{V}_p} \|z_i - a_i\|_2^2 \tag{17}$$

Combining these inequalities with $\mathcal{L}_{\mathrm{anch}} \leq \varepsilon_{\mathrm{A}}$ gives equation 13 and equation 14.

For the final result, define

$$\mathcal{S}_r = \{i \in \mathcal{V}_p : \|z_i - a_i\|_2 \geq r\} \tag{18}$$

Every $i \in \mathcal{S}_r$ contributes at least r^2 to the squared structural displacement. Hence,

$$|\mathcal{S}_r| r^2 \leq \sum_{i \in \mathcal{V}_p} \|z_i - a_i\|_2^2 \leq \frac{2n\varepsilon_{\mathrm{A}}}{\eta\beta_Z} \tag{19}$$

Dividing by nr^2 and using $|\mathcal{S}_r|/n \leq 1$ proves equation 15. □

Implication. Proposition 1 characterizes the effect of the deterministic manifold-anchoring regularizer. A small anchoring loss controls both the average magnitude of the sequence latent points and the average displacement of the structural latent points from their geometric anchors. The accompanying tail bound further limits the fraction of designed residues that can exhibit large structural deviations.

C.4 Sequence and Structural Reconstruction Objectives

The sequence and structural reconstruction terms used in equation 4 are defined as follows:

$$\mathcal{L}_{\mathrm{seq}} = -\frac{1}{|\mathcal{V}_p|} \sum_{i \in \mathcal{V}_p} \log p_\theta\left(s_i \mid Z_p, \mathcal{G}_r\right), \tag{20}$$

$$\begin{aligned}
\mathcal{L}_{\mathrm{str}} = {} & \lambda_X \frac{\sum_{i \in \mathcal{V}_p} \sum_{c=1}^{C} m_{i,c} \left\| \widehat{X}_{i,c} - X_{i,c} \right\|_2^2}{\sum_{i \in \mathcal{V}_p} \sum_{c=1}^{C} m_{i,c}} + \lambda_{C_\alpha} \frac{\sum_{i \in \mathcal{V}_p} m_{i,C_\alpha} \left\| \widehat{X}_{i,C_\alpha} - X_{i,C_\alpha} \right\|_2^2}{\sum_{i \in \mathcal{V}_p} m_{i,C_\alpha}} \\
& + \lambda_{\mathrm{bb}} \frac{1}{|\mathcal{B}_{\mathrm{bb}}|} \sum_{(a,b) \in \mathcal{B}_{\mathrm{bb}}} \left| d_{ab}(\widehat{X}) - d_{ab}(X) \right| + \lambda_{\mathrm{sc}} \frac{1}{|\mathcal{B}_{\mathrm{sc}}|} \sum_{(a,b) \in \mathcal{B}_{\mathrm{sc}}} \left| d_{ab}(\widehat{X}) - d_{ab}(X) \right| \\
& + \lambda_{\mathcal{Q}} \frac{1}{|\mathcal{T}_{\mathcal{Q}}|} \sum_{(a,b,c,d) \in \mathcal{T}_{\mathcal{Q}}} \Big[\left(\cos \Omega_{abcd}(\widehat{X}) - \cos \Omega_{abcd}(X) \right)^2 \\
& \qquad + \left(\sin \Omega_{abcd}(\widehat{X}) - \sin \Omega_{abcd}(X) \right)^2 \Big]
\end{aligned} \tag{21}$$

where $\mathcal{B}_{\mathrm{bb}}$ and $\mathcal{B}_{\mathrm{sc}}$ denote the sets of backbone and side-chain bonds within the peptide region, respectively, and $\mathcal{T}_{\mathcal{Q}}$ denotes the set of atom quadruplets used to define side-chain torsions. For an atom pair a, b, $d_{ab}(X) = \|X_a - X_b\|_2$ denotes the bond length induced by X, while Ω_{abcd} denotes the side-chain dihedral angle determined by the coordinates of atom quadruplet (a, b, c, d).

D Details of Latent Representation Alignment

D.1 Statistical Normalization of the Sequence Latent Variables

The standard latent diffusion objective requires the denoising network to recover the noise added to a perturbed state, but does not explicitly constrain its internal representations to remain consistent with the deterministic latent space constructed by MHAE.

Because deterministic projection is used in MHAE instead of an isotropic Gaussian posterior, the noise introduced by reparameterized sampling is avoided. However, different latent dimensions may still have distinct scales and offsets. If the latent diffusion model is trained directly on the original u_i, the denoiser must account simultaneously for global latent bias, dimension-wise scale variation, and diffusion noise. This additional burden may limit its ability to recover the original residue representations. LRA therefore estimates dimension-wise statistics from the peptide latent variables in the training set after MHAE has been trained. Let the training set $\mathcal{D}_{\text{train}} = \left\{\mathcal{G}^{(n)}\right\}_{n=1}^{N_{\text{train}}}$ contain N_{train} complexes. The sequence latent points of all peptide residues to be generated are collected as $\mathcal{U}_{\text{train}} = \left\{u_i^{(n)} \mid i \in \mathcal{V}_p^{(n)}, \mathcal{G}^{(n)} \in \mathcal{D}_{\text{train}}\right\}$. The dimension-wise mean and standard deviation are then defined as follows:

$$\begin{aligned}
\mu_H &= \frac{1}{\sum_{\mathcal{G}^{(n)} \in \mathcal{D}_{\text{train}}} |\mathcal{V}_p^{(n)}|} \sum_{\mathcal{G}^{(n)} \in \mathcal{D}_{\text{train}}} \sum_{i \in \mathcal{V}_p^{(n)}} u_i^{(n)}, \\
\sigma_H &= \sqrt{\frac{1}{\sum_{\mathcal{G}^{(n)} \in \mathcal{D}_{\text{train}}} |\mathcal{V}_p^{(n)}|} \sum_{\mathcal{G}^{(n)} \in \mathcal{D}_{\text{train}}} \sum_{i \in \mathcal{V}_p^{(n)}} \left(u_i^{(n)} - \mu_H\right) \odot \left(u_i^{(n)} - \mu_H\right)}
\end{aligned} \tag{22}$$

where $\odot$ denotes element-wise multiplication. To avoid numerical instability in dimensions with small variance, the lower-bounded standard deviation is defined as $\overline{\sigma}_{H,k} = \max\left(\sigma_{H,k}, \delta_H\right), k = 1, \ldots, d_H$, and $D_H = \operatorname{diag}\left(\overline{\sigma}_H\right)$, where $\delta_H > 0$ is a small stability constant and $\overline{\sigma}_H$ is the resulting lower-bounded standard-deviation vector. The MHAE output u_i is then normalized as $\widetilde{u}_i = D_H^{-1}\left(u_i - \mu_H\right)$. The latent diffusion model is therefore trained on a centered sequence-latent distribution with comparable scales across dimensions. During sampling, the generated latent point $\widetilde{u}_i$ is mapped back to the original latent space constructed by MHAE through $u_i = D_H \widetilde{u}_i + \mu_H$. This transformation aligns the scale of the sequence latent variables with the Gaussian prior used by the diffusion model and avoids inconsistencies between training and sampling caused by dimension-wise distribution shifts.

Proposition 2. Centering, variance control, and stability of LRA normalization Let $\mathcal{I}_{\text{train}}$ index all peptide-residue instances in the training set, and let $N = |\mathcal{I}_{\text{train}}|$. For the statistics defined in equation 22, consider the normalization map

$$\mathcal{N}_H(u) = D_H^{-1}(u - \mu_H), \qquad D_H = \operatorname{diag}(\overline{\sigma}_H) \tag{23}$$

where $\overline{\sigma}_{H,k} = \max(\sigma_{H,k}, \delta_H)$ and $\delta_H > 0$. Then the normalized training latents satisfy

$$\frac{1}{N} \sum_{r \in \mathcal{I}_{\text{train}}} \mathcal{N}_H(u_r) = 0 \tag{24}$$

and, for every latent dimension k,

$$\frac{1}{N} \sum_{r \in \mathcal{I}_{\text{train}}} [\mathcal{N}_H(u_r)]_k^2 = \left(\frac{\sigma_{H,k}}{\overline{\sigma}_{H,k}}\right)^2 \leq 1 \tag{25}$$

Equality holds in equation 25 whenever $\sigma_{H,k} \geq \delta_H$. Furthermore, for any two sequence latent points u and v,

$$\|\mathcal{N}_H(u) - \mathcal{N}_H(v)\|_2 \leq \frac{1}{\delta_H} \|u - v\|_2 \tag{26}$$

Proof. By the definition of μ_H,

$$\frac{1}{N}\sum_{r\in\mathcal{I}_{\text{train}}}(u_r-\mu_H)=0 \tag{27}$$

Multiplication by the fixed diagonal matrix D_H^{-1} therefore gives equation 24.

For dimension k, the definition of $\sigma_{H,k}$ gives

$$\frac{1}{N}\sum_{r\in\mathcal{I}_{\text{train}}}(u_{r,k}-\mu_{H,k})^2=\sigma_{H,k}^2 \tag{28}$$

Consequently,

$$\frac{1}{N}\sum_{r\in\mathcal{I}_{\text{train}}}\left(\frac{u_{r,k}-\mu_{H,k}}{\overline{\sigma}_{H,k}}\right)^2=\frac{\sigma_{H,k}^2}{\overline{\sigma}_{H,k}^2}\le 1 \tag{29}$$

because $\overline{\sigma}_{H,k}\ge\sigma_{H,k}$. If $\sigma_{H,k}\ge\delta_H$, then $\overline{\sigma}_{H,k}=\sigma_{H,k}$ and equality holds.

Finally,

$$\|\mathcal{N}_H(u)-\mathcal{N}_H(v)\|_2=\left\|D_H^{-1}(u-v)\right\|_2 \tag{30}$$

$$\le\left\|D_H^{-1}\right\|_2\|u-v\|_2 \tag{31}$$

Since every diagonal entry of D_H is at least δ_H, $\|D_H^{-1}\|_2\le\delta_H^{-1}$, proving equation 26. □

Implication. Proposition 2 shows that LRA normalization removes the empirical mean, ensures that no latent dimension has an empirical variance greater than one, and limits the amplification of latent perturbations. This result does not require the normalized latent distribution to be Gaussian or its dimensions to be statistically independent.

D.2 Forward Perturbation of the Normalized Sequence Latent Points

The diffusion model is therefore trained to map noisy states back to latent variables while progressively approaching the latent manifold. Based on this, $\widetilde{u}_i$ is used by LRA as the diffusion target. For a global diffusion step t, let $\overline{\alpha}_t=\prod_{\tau=1}^{t}\alpha_\tau$, where $\alpha_t=1-\beta_t$ and β_t controls the noise intensity at the t-th step of the forward diffusion process. The normalized sequence latent point is perturbed as follows:

$$\widetilde{u}_i^t=\sqrt{\overline{\alpha}_t}\,\widetilde{u}_i+\sqrt{1-\overline{\alpha}_t}\,\epsilon_{i,H}^t,\qquad \epsilon_{i,H}^t\sim\mathcal{N}\left(0,I_{d_H}\right) \tag{32}$$

LRA applies the standard diffusion schedule only to the sequence latent points, while the structural latent coordinates are perturbed according to the interface-adaptive schedule introduced by IGAD.

E Details of Interface-Guided Adaptive Diffusion

E.1 Training and Sampling of the Interface Anchoring Score

The following details specify how the interface anchoring score in equation 6 is used during training and sampling.

During training, χ_i^0 is computed from the clean structural latent state before forward noising and is held fixed when constructing $\widetilde{z}_i^t$. During reverse sampling, χ_i^t is recomputed from the current structural latent state. Residues closer to the receptor are more strongly anchored, while residues outside the generated peptide region have zero anchoring scores. Thus, IGAD adjusts only the peptide structural diffusion schedule, while the receptor condition remains unchanged.

E.2 Interpretation of the Residue-Adaptive Schedule

This uniform schedule does not account for the spatial heterogeneity of peptide binding interfaces. Residues near the receptor surface often determine the binding pose and key interactions and may therefore need to stabilize earlier during generation, whereas residues farther from the interface can retain greater conformational freedom.

For a residue far from the receptor interface, $\chi_i \to 0$ and $\mathcal{K}_i \to 1$, reducing the local diffusion process to the standard uniform schedule. For a residue close to the interface, $\chi_i \to 1$ and $\mathcal{K}_i \to \mathcal{K}_{\min} < 1$, allowing more unperturbed signal to be retained at the same global time step. Residues near the interface therefore retain a larger fraction of the unperturbed latent signal and receive less noise at the same global time step t. Without increasing the total number of diffusion steps, a locally compressed denoising trajectory is induced for these residues, allowing their conformations to stabilize earlier.

This design is motivated by the fact that receptor-peptide binding depends primarily on three-dimensional geometric complementarity and local conformational stability at the interface. Interface priors can therefore be incorporated more directly as generative constraints by modulating the structural latent space.

Proposition 3. Monotone interface prioritization and boundary preservation Fix a diffusion state s and let $\chi_i^s \in [0,1]$ be the interface anchoring score of peptide residue i. Define

$$\mathcal{K}_i = 1 - \chi_i^s(1 - \mathcal{K}_{\min}), \qquad 0 < \mathcal{K}_{\min} \leq 1 \tag{33}$$

and, for $0 < \overline{\alpha}_t < 1$,

$$\overline{\alpha}_{t,i}^{\mathrm{IGAD}} = (\overline{\alpha}_t)^{\mathcal{K}_i} \tag{34}$$

Then

$$\mathcal{K}_{\min} \leq \mathcal{K}_i \leq 1 \tag{35}$$

and

$$\overline{\alpha}_t \leq \overline{\alpha}_{t,i}^{\mathrm{IGAD}} \leq (\overline{\alpha}_t)^{\mathcal{K}_{\min}} \tag{36}$$

For any two peptide residues i and j,

$$\chi_i^s \geq \chi_j^s \quad \Longrightarrow \quad \overline{\alpha}_{t,i}^{\mathrm{IGAD}} \geq \overline{\alpha}_{t,j}^{\mathrm{IGAD}} \tag{37}$$

and therefore

$$1 - \overline{\alpha}_{t,i}^{\mathrm{IGAD}} \leq 1 - \overline{\alpha}_{t,j}^{\mathrm{IGAD}} \tag{38}$$

The corresponding residue-level noise intensities also satisfy

$$\chi_i^s \geq \chi_j^s \quad \Longrightarrow \quad \beta_{t,i}^{\mathrm{IGAD}} \leq \beta_{t,j}^{\mathrm{IGAD}} \tag{39}$$

At the diffusion boundaries,

$$\overline{\alpha}_{0,i}^{\mathrm{IGAD}} = 1, \qquad \sqrt{\overline{\alpha}_{T,i}^{\mathrm{IGAD}}} \leq (\overline{\alpha}_T)^{\mathcal{K}_{\min}/2} \tag{40}$$

Thus, if $(\overline{\alpha}_T)^{\mathcal{K}_{\min}} \ll 1$, the terminal residual signal is uniformly negligible for all peptide residues.

Proof. Because $\chi_i^s \in [0,1]$, the affine map

$$\mathcal{K}_i = 1 - \chi_i^s(1 - \mathcal{K}_{\min}) \tag{41}$$

maps the interval $[0,1]$ to $[\mathcal{K}_{\min}, 1]$, proving equation 35.

For any fixed $x \in (0,1)$, the function $f(k) = x^k$ is decreasing in k, since

$$\frac{\partial x^k}{\partial k} = x^k \log x < 0 \tag{42}$$

Applying this result with $x = \overline{\alpha}_t$ and using $\mathcal{K}_i \in [\mathcal{K}_{\min}, 1]$ yields equation 36.

Moreover, $\chi_i^s \geq \chi_j^s$ implies $\mathcal{K}_i \leq \mathcal{K}_j$. Because $(\overline{\alpha}_t)^k$ decreases with k, this gives equation 37. Subtracting both sides from one gives equation 38.

Similarly, since

$$\beta_{t,i}^{\mathrm{IGAD}} = 1 - (1 - \beta_t)^{\mathcal{K}_i} = 1 - \alpha_t^{\mathcal{K}_i} \tag{43}$$

a smaller $\mathcal{K}_i$ produces a larger $\alpha_t^{\mathcal{K}_i}$ and therefore a smaller $\beta_{t,i}^{\mathrm{IGAD}}$. This proves equation 39.

Finally, $\overline{\alpha}_0 = 1$ gives $\overline{\alpha}_{0,i}^{\mathrm{IGAD}} = 1$ for every $\mathcal{K}_i$. At $t = T$, equation 36 gives

$$\overline{\alpha}_{T,i}^{\mathrm{IGAD}} \leq (\overline{\alpha}_T)^{\mathcal{K}_{\min}} \tag{44}$$

Taking square roots proves equation 40. □

Implication and scope. Proposition 3 establishes that, at the same global step, residues with larger interface anchoring scores retain at least as much structural signal and receive no more diffusion noise than residues with smaller scores. It also shows that IGAD preserves the clean-state boundary and provides a uniform bound on the terminal residual signal. During state-adaptive reverse sampling, these guarantees hold conditionally at each score update given the current structural state.

E.3 Residue-Level Time Conditioning

Standard diffusion models typically provide the global β_t to the denoising network as a shared time signal. Under IGAD, this global signal is extended to the residue-level local time signal $\beta_{t,i}^{\mathrm{IGAD}} = 1-(1-\beta_t)^{\mathcal{K}_i}$. Based on this local noise intensity, the residue-level time embedding is defined as $\gamma_i^{\mathrm{IGAD}}(t) = \left[\beta_{t,i}^{\mathrm{IGAD}}, \sin\left(\beta_{t,i}^{\mathrm{IGAD}}\right), \cos\left(\beta_{t,i}^{\mathrm{IGAD}}\right)\right]$. The denoiser is thus conditioned on spatially varying diffusion times determined by interface distance rather than on a single global time signal:

$$\left(\widehat{\epsilon}_H^t, \widehat{\epsilon}_Z^t, R_\theta^t\right) = \epsilon_\theta\Big(\widetilde{H}^t, \widetilde{X}^t, \{\gamma_i^{\mathrm{IGAD}}(t)\}_{i\in\mathcal{V}}, \{\mathrm{Emb}_p(p_i)\}_{i\in\mathcal{V}}, \mathcal{E}_{\mathrm{ctx}}, \mathcal{E}_{\mathrm{int}}\Big) \tag{45}$$

where ϵ_θ denotes the parameterized denoising network, $\widehat{\epsilon}_H^t$ and $\widehat{\epsilon}_Z^t$ denote the predicted noise for sequence and structural latent coordinates, respectively, and R_θ^t denotes the residue-level representation produced by an intermediate denoiser layer. This design allows the denoising network to account for the different local diffusion stages of individual residues. Residues near the interface remain closer to the latent manifold and are stabilized earlier, whereas residues farther from the interface retain greater freedom for exploration. Interface priority is therefore encoded directly into the diffusion time rather than supplied only as a static conditioning signal.

E.4 State-Adaptive Reverse Sampling

During reverse sampling, IGAD dynamically updates the interface anchoring scores based on the current noisy structural latent state $\widetilde{z}^t$. Specifically, χ_i^t and $\mathcal{K}_i^t$ are recomputed every K sampling steps. At each reverse step, the current $\mathcal{K}_i^t$ is held fixed and used to define the step-local coefficients $\alpha_{t,i}^{\mathrm{IGAD}} = (\alpha_t)^{\mathcal{K}_i^t}$, $\quad \overline{\alpha}_{t,i}^{\mathrm{IGAD}} = (\overline{\alpha}_t)^{\mathcal{K}_i^t}$, $\quad \overline{\alpha}_{t-1,i}^{\mathrm{IGAD}} = (\overline{\alpha}_{t-1})^{\mathcal{K}_i^t}$ The corresponding transition variance is

$$\left(\sigma_{t,i}^{\mathrm{IGAD}}\right)^2 = \frac{\left(1-\alpha_{t,i}^{\mathrm{IGAD}}\right)\left(1-\overline{\alpha}_{t-1,i}^{\mathrm{IGAD}}\right)}{1-\overline{\alpha}_{t,i}^{\mathrm{IGAD}}} \tag{46}$$

These coefficients define a state-adaptive local reverse update rather than the exact posterior of a globally fixed diffusion chain. The structural latent state is then updated through the reverse process as follows:

$$\widetilde{z}_i^{t-1} = \frac{1}{\sqrt{\alpha_{t,i}^{\mathrm{IGAD}}}}\left(\widetilde{z}_i^t - \frac{1-\alpha_{t,i}^{\mathrm{IGAD}}}{\sqrt{1-\overline{\alpha}_{t,i}^{\mathrm{IGAD}}}}\widehat{\epsilon}_{i,Z}^t\right) + \sigma_{t,i}^{\mathrm{IGAD}}\xi_{i,Z}^t, \qquad \xi_{i,Z}^t \sim \mathcal{N}\left(0, I_{d_Z}\right) \tag{47}$$

In contrast, the sequence latent state does not use the interface-adaptive exponent $\mathcal{K}_i^t$ and follows the standard diffusion schedule during the reverse update:

$$\widetilde{u}_i^{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\widetilde{u}_i^t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\widehat{\epsilon}_{i,H}^t\right) + \sqrt{\frac{(1-\alpha_t)(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}}\xi_{i,H}^t, \qquad \xi_{i,H}^t \sim \mathcal{N}\left(0, I_{d_H}\right) \tag{48}$$

IGAD therefore does not impose an additional heuristic constraint outside the diffusion trajectory. Instead, interface priority is incorporated directly into the diffusion Markov chain.

E.5 Latent Diffusion Objective and Boundary Conditions

For each training sample, a diffusion step is randomly drawn as $t \sim \mathrm{Uniform}(1, \ldots, T)$. The structural noise-prediction loss is computed only over the peptide region under design and its valid atom

channels, whereas the sequence noise-prediction loss is computed over the normalized sequence latent variables. The latent diffusion objective further includes LRA. The final latent diffusion objective of PepMALD is defined as follows:

$$\mathcal{L}_{\mathrm{LDM}} = \mathbb{E}_{t,\mathcal{G},\epsilon_H,\epsilon_Z}\Bigg[\omega_H \frac{1}{|\mathcal{V}_p|}\sum_{i\in\mathcal{V}_p}\left\|\epsilon_{i,H}^t - \widehat{\epsilon}_{i,H}^t\right\|_2^2 + \frac{\sum_{i\in\mathcal{V}_p}\sum_{c=1}^{C} m_{i,c}\left\|\epsilon_{i,Z}^t - \widehat{\epsilon}_{i,Z}^t\right\|_2^2}{\sum_{i\in\mathcal{V}_p}\sum_{c=1}^{C} m_{i,c}} + \lambda_R\mathcal{L}_{\mathrm{LRA}}\Bigg] \tag{49}$$

where the structural perturbation associated with $\epsilon_{i,Z}^t$ is scaled by $\overline{\alpha}_{t,i}^{\mathrm{IGAD}}$ in equation 8. Thus, the structural noise-regression term is evaluated on latent states generated under the interface-adaptive schedule. The coefficient ω_H balances the sequence and structural noise-regression losses, while λ_R controls the contribution of the LRA term.

The ordering and boundary properties of the interface-adaptive schedule are established in Proposition 3. In particular, the clean-state boundary remains unchanged, while the terminal residual signal is uniformly bounded by $(\overline{\alpha}_T)^{\mathcal{K}_{\min}}$. Under the schedule used in this study, this upper bound is negligible. IGAD therefore preserves the global diffusion boundaries while introducing spatially varying signal-retention rates at intermediate steps.

F Experimental Setup and Extended Generalization Results

F.1 Dataset Construction and Leakage Control

Three datasets from different sources are used to evaluate PepMALD for peptide design. PepBench and PepBDB (Wen et al., 2019) contain complex structures with explicit receptor-peptide interactions and are used for supervised training and evaluation. ProtFrag is constructed by segmenting single-chain protein structures from the PDB (Berman et al., 2000) and provides a large collection of local protein fragments for learning general conformational patterns. The lengths of all peptides and fragments are restricted to 4-25 residues. The experimental setting of PepGLAD is followed while preventing information leakage at the receptor level. For PepBench, LNR (Tsaban et al., 2022) is used as the test set to evaluate generalization to unseen receptor targets. To prevent highly similar receptors from appearing in both the training data and the test set, the receptor sequences of LNR samples and candidate protein-peptide complexes from the PDB are jointly clustered. Receptor clusters are defined using MMseqs2 (Steinegger & Söding, 2017) with a sequence identity threshold of 40%. Any complex whose receptor belongs to the same cluster as an LNR receptor is removed from the candidate training and validation sets. The remaining samples are then split into training and validation sets at the cluster level. Redundant samples with excessively high receptor or peptide sequence identity are further removed from the supervised data. For PepBDB, a nonredundant split is constructed using MMseqs2 clustering to evaluate model generalization and robustness on an independently collected set of protein-peptide complex structures.

F.2 Baseline Selection and Reproduction Protocol

Receptor-conditioned full-atom peptide design requires the joint modeling of sequence, structure, and interface constraints, but few existing methods directly address this setting. PepMALD is compared with two groups of representative baselines. The first group includes HSRN (Jin et al., 2022) and dyMEAN (Kong et al., 2023), which support receptor-conditioned full-atom structure generation. Although these methods were developed primarily for antibody design, both can generate full-atom structures conditioned on a receptor and thus provide useful references for evaluating how well different full-atom generation methods transfer to peptide design. The second group includes the peptide design methods PepGLAD (Kong et al., 2024), PepMimic (Kong et al., 2025), and PepTri (Nguyen et al., 2026), which represent geometric latent diffusion, interface mimicry, and peptide generation guided by multiple constraints, respectively. For a fair comparison, all methods

are reproduced using unified data processing, sampling, and evaluation protocols, while the training configurations reported in their original papers are followed.

F.3 Evaluation Metrics

Model performance is evaluated from several perspectives. Receptor-peptide interaction quality is assessed using the Rosetta (Alford et al., 2017) binding energy ΔG, which reflects the ability of a generated peptide to form stable interactions with the receptor. DockQ is used to measure interface-level similarity between the generated and reference complexes. Structural accuracy is evaluated using RMSD, which quantifies the spatial deviation of a generated conformation from the reference structure. Overall candidate quality is assessed using Success, defined as the proportion of samples that meet the predefined criteria for structural validity and energetic quality. The properties of the generated distribution are also evaluated. Diversity (Div) measures the variation among candidates generated for the same receptor in sequence and structural spaces, whereas Consistency (Con) assesses whether sequence and structural differences change in a coordinated manner.

F.4 Extended Analysis of Sequence-Structure Co-Design

Although several latent diffusion baselines perform well on individual metrics, these gains do not consistently yield more usable candidates. Greater diversity or sequence variation alone is therefore insufficient to ensure valid binding conformations. PepTri achieves higher Diversity on both datasets, but this broader distribution does not produce corresponding improvements in usable designs. On PepBDB, it is also accompanied by a more pronounced decline in structural quality. A broader generative distribution therefore does not necessarily improve candidate quality in receptor-conditioned peptide design. Without adequate constraints on feasible conformations and interface geometry, greater diversity may increase the number of invalid samples.

F.5 Additional Analysis of Binding Conformation Generation

Once the sequence is fixed, improvements depend mainly on modeling the conformational space and imposing receptor-interface constraints rather than optimizing the sequence. The results on both datasets suggest that manifold-anchored latent representations provide a stable basis for structural sampling, while diffusion guided by interface information reduces invalid conformations that deviate from the receptor binding region.

F.6 Construction and Characteristics of the Post-2026 Test Set

The established PepBench and PepBDB splits provide standardized protocols for evaluating receptor-conditioned peptide generation. However, these fixed benchmarks do not directly measure whether a model can generalize to receptor-peptide complexes released after the training-data cutoff, especially when related receptor structures may already exist in the training-related datasets. We therefore construct the Post-2026 test set consisting of newly released complexes with low receptor sequence similarity to those in the training-related datasets.

The Post-2026 set includes complexes released after January 1, 2026. To minimize receptor-level redundancy with the training datasets, we retain only complexes with receptor sequence identity below 0.40. As shown in Figure 6A, the resulting benchmark contains 410 receptor-peptide complexes from 189 unique PDB entries. At the PDB-entry level, this set is larger than LNR, which contains 93 entries, and similar in size to the PepBDB test set, which contains 190 entries. These data establish a post-cutoff benchmark for assessing model generalization to newly released complexes. Its dataset characteristics are summarized in Figure 6B. The peptide length median in the Post-2026 set is 14 amino acids, with an interquartile range of 8-19 amino acids. By comparison, LNR has a median peptide length of 10 amino acids and an interquartile range of 6-14 amino acids. The Post-2026 set therefore contains a larger proportion of medium and long peptides. As shown in Figure 6C, its peptide sequence entropy is 2.77, which is also higher than that of LNR, while its aromatic residue fraction is 0.103. These properties provide an evaluation setting whose sequence composition differs from that of the standard LNR split.

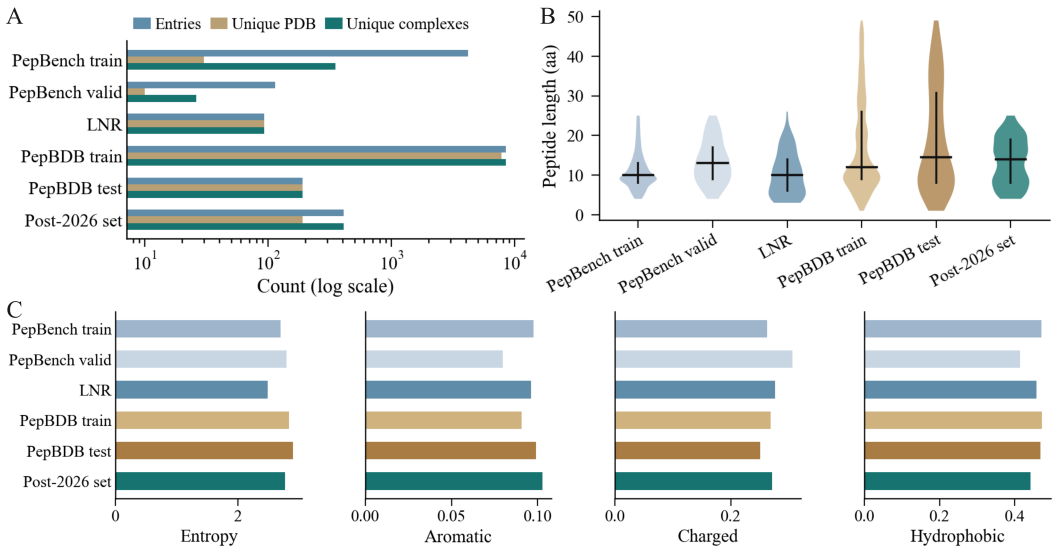

Figure 6: **Construction and characteristics of the Post-2026 test set. A.** Scale and nonredundancy of the Post-2026 set. **B.** Peptide length distribution across evaluation sets. **C.** Peptide sequence properties of the Post-2026 set.

G Additional Analysis of Manifold-Anchored Latent Space Quality

G.1 Encoding Stability and Latent-Manifold Structure

The residue-level latent variables produced by MHAE are unaffected by stochastic posterior sampling and provide stable, deterministic representations. This pattern is also observed in the cross-model analysis. As shown in Figure 7A, PepMALD exhibits near-zero encoding variability, whereas the VAE-based baselines show substantially greater variability due to posterior sampling.

The PCA projections further show that the residue-level latent variables of PepMALD form a more compact and continuous distribution. A clearer local organization of interface residues can also be observed in Figure 7B. Together, these results suggest that MHAE maps the residues under design onto a stable latent manifold rather than representing them through stochastic latent samples.

G.2 Dimensional Utilization

The cross-model comparison in Figure 7C further shows that the latent variance of PepMALD is distributed more evenly across dimensions, without being dominated by a small subset. Its normalized participation ratio reaches 0.856, exceeding those of the VAE-based baselines. MHAE therefore reduces VAE-style latent collapse and increases the information capacity of the residue-level representations.

G.3 Latent Prior Compatibility

The latent distribution should also be compatible with the Gaussian prior used during diffusion. When substantial scale shifts, interdimensional correlations, or prior mismatch are present, the denoiser must learn both distribution correction and noise regression, making generation more difficult. PepMALD obtains lower KS-to-normal scores and weaker interdimensional correlations than the VAE-based baselines, as shown in Figure 7D. The latent space constructed by MHAE is therefore closer to the standard Gaussian prior commonly used in diffusion models and contains less redundant dependence across dimensions. This provides a more stable distributional basis for latent normalization and subsequent latent diffusion training.

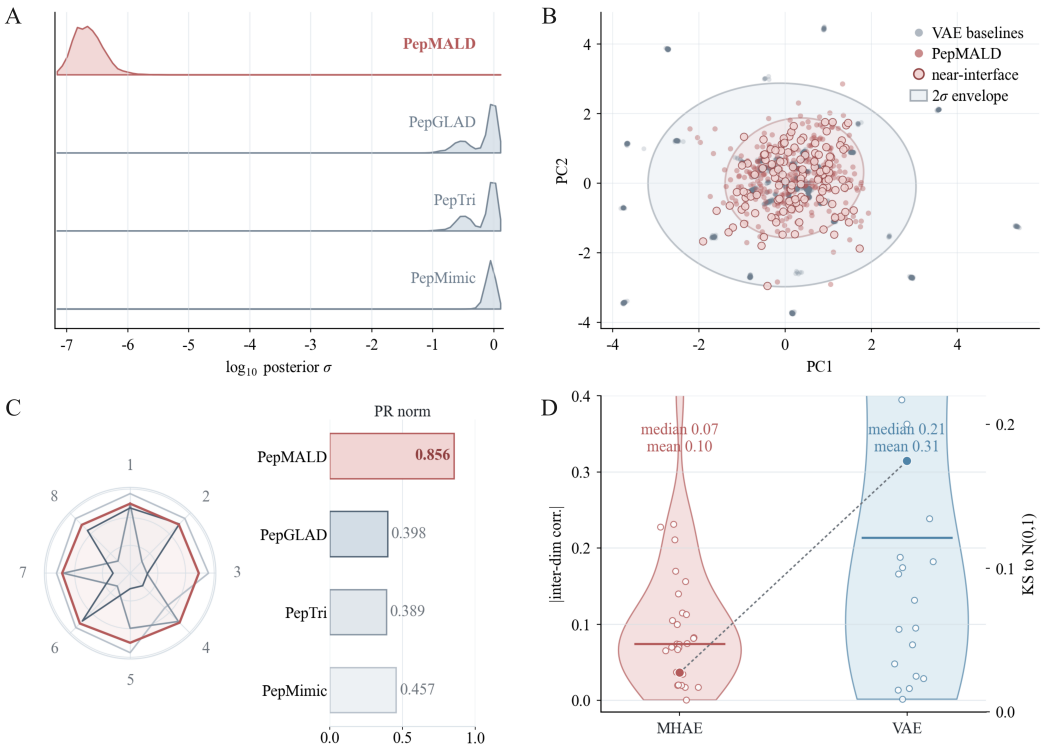

Figure 7: **A.** Distribution of encoding variability across latent dimensions and residues. **B.** PCA projection of residue-level latent variables from PepMALD and VAE-based baselines. PepMALD forms a compact manifold, with near-interface residues highlighted by outlined markers; ellipses indicate two-sigma envelopes of the latent distributions. **C.** Radar plot of per-dimension latent variance, showing more uniform utilization of latent dimensions by PepMALD. **D.** Latent prior compatibility. MHAE exhibits lower interdimensional correlation and smaller KS distance to the standard normal.

G.4 Frozen-Latent Linear-Probe Protocol

The interface information retained in the latent representations is evaluated using frozen-latent linear probing. The MHAE encoder is frozen, and a linear probe is trained using only the latent vector of each residue to predict two interface-related properties. Interface contact is defined according to whether the minimum residue-to-receptor C_α-C_α distance is below a contact threshold and is evaluated using logistic regression. Receptor distance is predicted using ridge regression. Because these low-capacity linear probes have no access to coordinates, receptor structures, or additional features, successful prediction indicates that interface information is linearly accessible from the frozen latent representations.

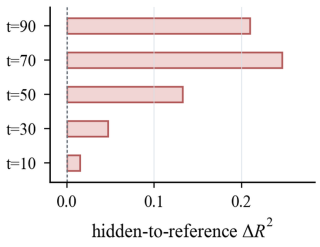

Figure 8: Step-wise gain in hidden-to-reference reconstruction R^2, showing that latent alignment provides the largest improvement during intermediate denoising.

H Additional Analysis of Latent Diffusion Dynamics

H.1 Step-Wise Effect of Latent Representation Alignment

The hidden-to-reference reconstruction R^2 measures how much reference latent information can be linearly recovered from the intermediate representation of the denoiser. Although LRA improves

this recoverability throughout reverse diffusion, the magnitude of the improvement varies across diffusion steps.

As shown in Figure 8, the step-wise gain is concentrated in the middle stages of diffusion and is most pronounced near $t = 70$ and $t = 50$. At these stages, the latent variables remain strongly corrupted, while sequence and structural information has begun to recover. The reverse trajectory is therefore particularly likely to deviate from the latent manifold. The larger R^2 gains in this interval indicate that the latent reference provides a stable internal reference for the denoiser, allowing the hidden states to approach the MHAE latent space earlier in the reverse process.

H.2 Interface-Dependent Reverse Transitions

We next examine whether IGAD changes the residue-level reverse transitions. A traditional LDM uses a globally shared diffusion schedule, so different residues generally receive the same transition intensity at a given time step. In contrast, IGAD adjusts the local transition coefficient according to the interface anchor score of each residue, producing different denoising dynamics for residues near and far from the receptor interface.

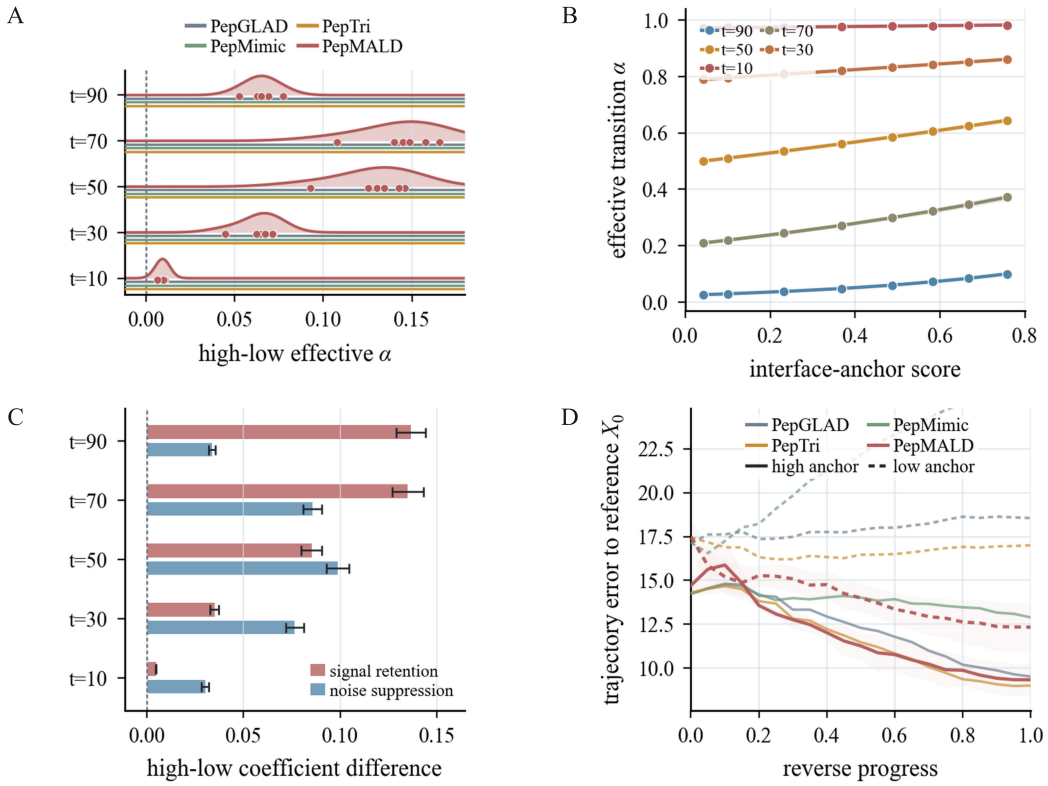

Figure 9: **A.** Ridge-density distributions of the effective-α gap between high- and low-anchor residues across diffusion steps. **B.** Relationship between residue anchor score and effective transition α at representative diffusion steps. **C.** Decomposition of the high-anchor advantage into signal retention and noise reduction across diffusion steps. **D.** Reverse-trajectory error to the reference latent state for high- and low-anchor residues.

Figure 9A shows the distribution of the effective-α gap between high- and low-anchor residues. For the baseline methods, this gap remains near zero across diffusion steps, indicating that their local transitions do not vary with the interface-anchor score. PepMALD exhibits nonzero high-low gaps at multiple diffusion steps, indicating different local reverse transitions for interface-related and distal residues.

Figure 9B further shows that the effective transition α increases across interface-anchor-score bins, particularly at intermediate diffusion steps. In contrast, the near-zero gaps observed for the baseline methods reflect their spatially uniform schedules. Thus, IGAD assigns residue-specific transition strengths according to interface proximity.

Figure 9C compares signal retention and noise reduction between high- and low-anchor residues. High-anchor residues retain more signal at $t = 90$ and $t = 70$, while their noise reduction is most evident around $t = 50$. These differences decrease toward $t = 10$, supporting earlier stabilization of interface-proximal residues while preserving the flexibility of distal residues.

This interface-prioritized local schedule is also reflected in the reverse-trajectory error. As shown in Figure 9D, high-anchor residues in PepMALD approach the reference latent state more rapidly than those in the baseline models. They also maintain lower trajectory errors near the end of reverse diffusion, whereas the errors of low-anchor residues decrease more gradually.

These complementary analyses explain the residue-level Markov heterogeneity reported in Figure 4B. IGAD does not impose a uniform denoising constraint on all residues. Instead, it adjusts the structural recovery process according to interface proximity without changing the global starting and ending points of diffusion.

I Case-Level Analysis of Interface-Aware Denoising Dynamics

We further examine whether the effects of manifold anchoring and residue-level transition heterogeneity observed at the aggregate level can also be identified within individual receptor-peptide complexes. If local reverse transitions are modulated according to interface anchoring, high-anchor residues should approach the reference latent state earlier than low-anchor residues. They should also exhibit larger effective transitions during critical intermediate denoising stages. To test this hypothesis, we analyze the reverse diffusion trajectories of PepMALD in both training-distribution and independent test cases. For each receptor-peptide complex, the interface-anchor score is calculated from the spatial relationship between each generated residue and the receptor pocket in the reference structure. The generated residues are then divided into high- and low-anchor groups. Along the reverse denoising trajectory, we record the trajectory error relative to the reference latent state, the effective transition α, and the anchor-α correlation. The $\mathrm{AUC}(\text{low error} - \text{high error})$ measures whether high-anchor residues consistently maintain lower trajectory errors during reverse generation. We also calculate the low-high trajectory-error gap, high-low effective-α gap, and anchor-α correlation at $t = 70$.

As shown in Figure 10A-B, interface-first denoising is observed in the majority of cases. Among the training-distribution cases, 78.9% have a positive $\mathrm{AUC}(\text{low error} - \text{high error})$, and 76.3% exhibit a positive low-high trajectory-error gap at $t = 70$. In the independent test cases, the corresponding proportions are 65.6% and 66.7%, respectively. These results indicate a general tendency for high-anchor residues to approach the reference latent state earlier, although the tendency is stronger in the training-distribution cases.

The positive proportions of the high-low effective-α gap and anchor-α correlation are more moderate than those of the trajectory-error measures. In the training-distribution cases, positive high-low effective-α gaps and anchor-α correlations are observed in 54.4% and 53.5% of the samples at $t = 70$, respectively. The corresponding proportions in the independent test cases are 52.7% and 53.8%. This behavior is expected because the dynamic anchor is continuously updated according to the current noisy structure during sampling. The effective transition α at a single time step therefore need not be perfectly monotonic with the groups defined by the interface-anchor score. Nevertheless, interface anchoring affects the cumulative trajectory more clearly, with high-anchor residues showing an earlier reduction in trajectory error relative to the reference latent state.

Figure 10C shows that the training-distribution cases 2KQS and 7NLC and the independent test cases 3WBN, 4RS9, 3R42, and 5LM1 lie in the region with a positive low-high trajectory-error gap. Among these cases, 3WBN exhibits both a clear low-high trajectory-error gap and a positive high-low effective-α gap.

The 3WBN complex contains 18 generated residues, including five high-anchor and five low-anchor residues. Its $\mathrm{AUC}(\text{low error} - \text{high error})$ is 11.71, and its low-high trajectory-error gap at $t = 70$

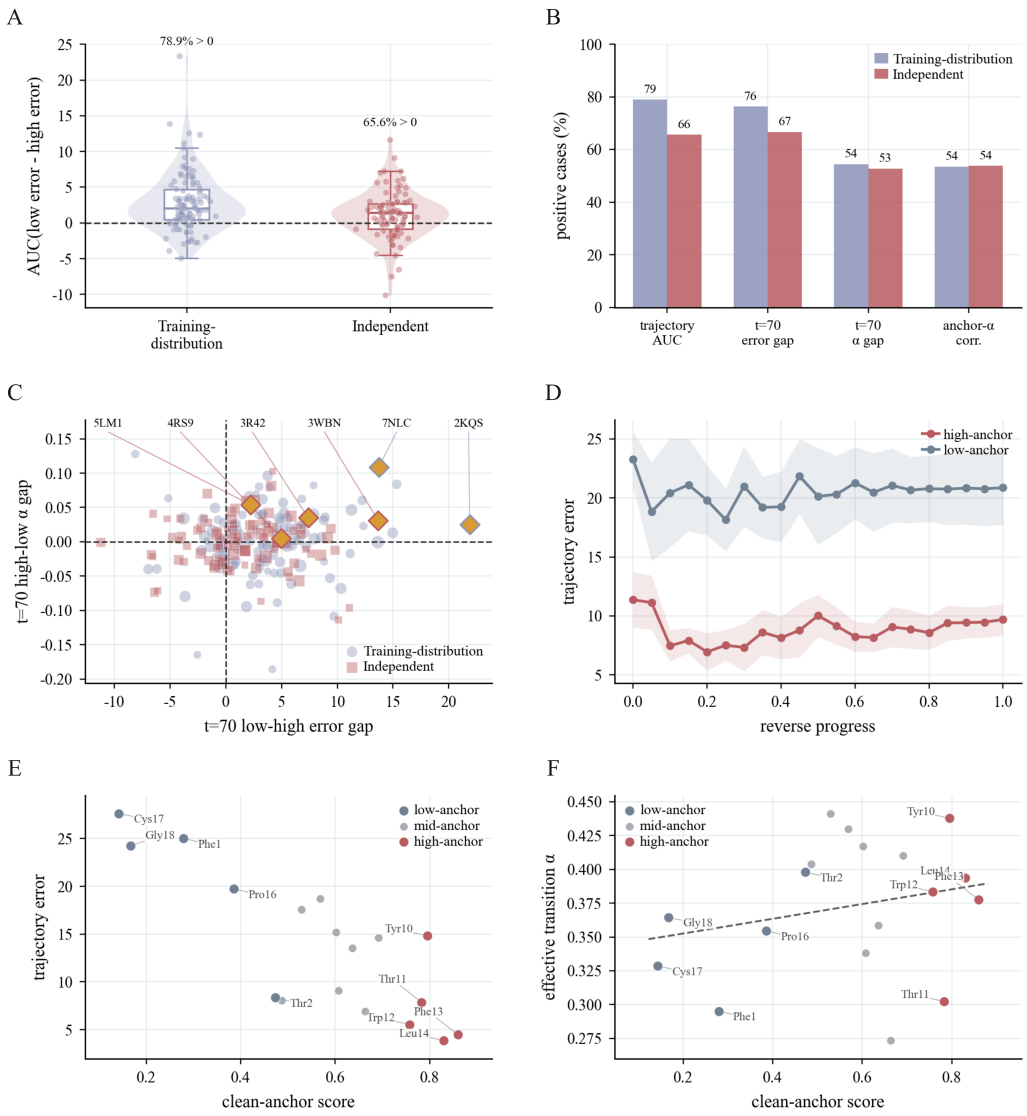

Figure 10: **A.** Distribution of trajectory separation in training-distribution and independent test cases, measured by the AUC of low-anchor minus high-anchor trajectory error. Positive values indicate that high-anchor residues approach the reference latent state earlier than low-anchor residues. **B.** Fraction of cases showing positive trajectory separation, positive trajectory-error gap at $t = 70$, positive effective-α gap, and positive anchor-α correlation. **C.** Representative-case selection map based on the low-high trajectory-error gap and high-low effective-α gap at $t = 70$. Highlighted structures denote the representative cases used for residue-level analysis. **D.** Reverse denoising trajectory of the independent 3WBN case, comparing high- and low-anchor residues. **E.** Residue-level relationship between interface-anchor score and trajectory error at $t = 70$ in 3WBN. **F.** Residue-level relationship between interface-anchor score and effective transition α at $t = 70$ in 3WBN. The dashed line denotes the linear trend across generated residues.

is 13.67. The trajectory-error gap remains 11.18 at the final stage. As shown in Figure 10D, the high-anchor residues maintain a lower trajectory error throughout reverse diffusion, indicating that residues near the receptor interface enter stable latent states earlier.

Figure 10E-F further explains the trajectory separation in 3WBN at the residue level. At $t = 70$, the high-anchor residues are Tyr10, Thr11, Trp12, Phe13, and Leu14, with a mean trajectory error of

7.324 and a mean effective transition α of 0.379. The low-anchor residues are Phe1, Thr2, Pro16, Cys17, and Gly18, with a mean trajectory error of 20.990 and a mean effective transition α of 0.348.

During intermediate denoising, the high-anchor residues are closer to the reference latent state and receive stronger local transitions. The sample-level and residue-level results provide additional evidence that PepMALD does not recover all residues through a globally uniform process. Instead, residue-level denoising priorities are adjusted according to receptor-peptide interface information, producing a spatially adaptive conditional generation trajectory.

J Case Study: Peptide Generation for Colorectal-Cancer-Relevant APC-β-Catenin Interfaces

J.1 Biological Setting and Target Selection

PepMALD was evaluated in a disease-relevant peptide-generation setting centered on the APC-β-catenin axis in colorectal cancer. APC is a key tumor suppressor in the Wnt signaling pathway, and APC mutations are strongly associated with familial adenomatous polyposis and sporadic colorectal tumors. β-catenin is regulated through direct binding to a series of 15- and 20-amino-acid repeats in APC, which contributes to its downregulation and degradation (Spink et al., 2001). These structurally characterized APC-derived binding motifs provide a biologically relevant setting for examining whether generated peptides can form plausible binding modes on disease-related protein interfaces.

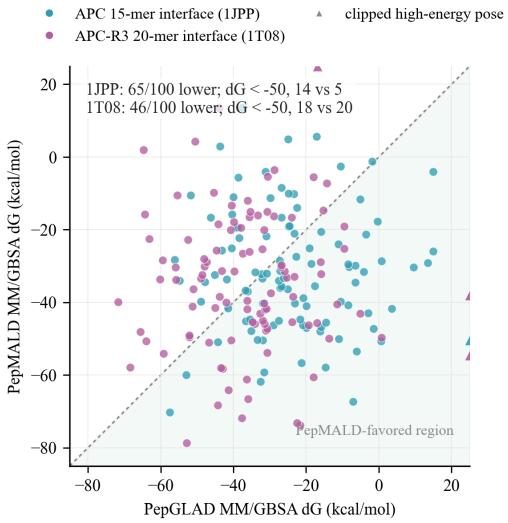

Figure 11: Candidate-level MM/GBSA comparison on two APC-β-catenin interfaces.

Two APC-related β-catenin interfaces were selected: the APC 15-mer repeat interface from PDB 1JPP and the APC-R3 20-mer repeat interface from PDB 1T08. APC-derived peptides bind the armadillo-repeat domain of β-catenin, which forms an extended groove for peptide recognition. The 1JPP interface provides a compact binding case involving acidic, hydrophobic, and polar interactions. APC-R3 represents a longer 20-mer interface, where phosphorylation-dependent interactions can expand the binding region and strengthen β-catenin recognition (Ha et al., 2004). These structures allow compact and extended APC-related binding grooves to be evaluated in parallel.

J.2 Candidate-Level Energy Evaluation

For each interface, PepMALD and PepGLAD generated 100 candidate peptide-β-catenin complexes. Every candidate was evaluated using MM/GBSA binding energy and OPLS (Optimized Potentials for Liquid Simulations) potential energy. MM/GBSA measures the relative binding favorability of the generated peptide at the β-catenin interface, whereas OPLS potential energy reflects the energetic state of the generated complex conformation.

As shown in Figure 11, PepMALD showed a clear MM/GBSA advantage on the compact 1JPP interface. Among the 100 paired candidates, 65 had lower binding energies for PepMALD than for PepGLAD, and PepMALD produced more strong binders. On the extended 1T08 interface, the two distributions were closer. PepMALD achieved lower binding energy in 46 of 100 comparisons and produced a comparable number of strong binders, indicating that binding-energy ranking on this interface was more dependent on the individual conformation.

OPLS potential energy showed a more consistent pattern across the two targets. As shown in Figure 12, PepMALD achieved lower potential energy than PepGLAD for 72 of 100 candidates on 1JPP and 78 of 100 candidates on 1T08.

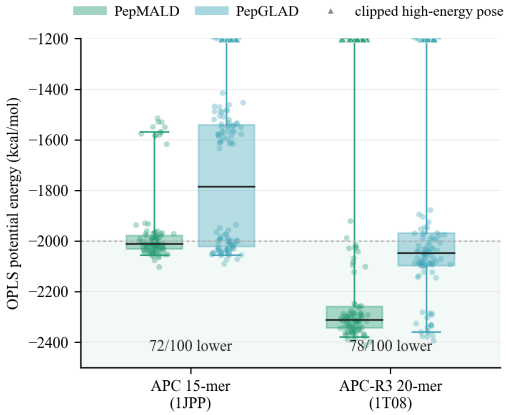

Figure 12: OPLS potential energy distributions of PepMALD and PepGLAD candidates on the APC 15-mer and APC-R3 20-mer interfaces.

J.3 Representative Binding Modes and Local Interactions

As shown in Figure 5A, at the compact 1JPP interface, the representative PepMALD peptide achieved an MM/GBSA binding energy of -67.36 kcal/mol, 9.82 kcal/mol lower than that of the corresponding PepGLAD peptide. The PepMALD pose followed the local groove geometry and contacted both hydrophilic and hydrophobic surface regions. By contrast, the PepGLAD peptide occupied a smaller region with weaker shape complementarity. At the extended 1T08 interface, the representative PepMALD and PepGLAD peptides achieved binding energies of -73.81 and -68.33 kcal/mol, respectively. The PepMALD pose extended across a longer region of the groove, consistent with APC 20-mer recognition, whereas the PepGLAD pose remained more localized. Across both interfaces, the lower predicted energies of the representative PepMALD peptides coincided with more extensive engagement of the β-catenin groove.

As shown in Figure 5B, the local interaction patterns provide additional detail. On 1JPP, the PepMALD peptide formed a distributed interaction network involving ASN236, SER238, ASP239, ASN275, ASN279, SER322, ARG323, and ASN365. These contacts extended along the peptide-binding path and provided multiple receptor-facing anchors. The PepGLAD peptide interacted mainly with ASN236, SER274, ASN279, and GLU311, resulting in a more localized contact pattern. A similar distinction was observed on 1T08. The PepMALD peptide contacted ASN284, ARG328, GLU413, and GLU416 across different regions of the groove. Although the PepGLAD peptide contacted ASN241, ASN284, LYS289, HIS324, and SER327, these interactions were concentrated around a local peptide segment. Thus, the difference between the models reflected the spatial organization of the contact network rather than contact number alone, with PepMALD achieving broader predicted groove engagement across both interfaces.

	Div	Con	ΔG	Success	DockQ	RMSD	Avg
PepMALD	**0.492**	**0.804**	**-22.72**	**0.611**	**0.689**	**2.84**	**0.654**
w/o MHAE	0.553	0.832	-18.17	0.542	0.672	3.27	**0.627**
w/o Global	0.379	0.776	-19.75	0.602	0.704	2.73	**0.620**
w/o Agg	0.416	0.792	-18.01	0.553	0.717	2.83	**0.614**
w/o MALD	0.493	0.805	-18.81	0.611	0.672	3.06	**0.631**
w/o LRA	0.467	0.803	-19.54	0.609	0.698	2.92	**0.636**
w/o IGAD	0.513	0.810	-20.02	0.506	0.684	3.04	**0.625**
w/o SN	0.475	0.845	-18.49	0.548	0.699	2.98	**0.628**

Figure 13: Ablation analysis of PepMALD components on PepBench. The Avg score is computed through (Div+Con+(-ΔG/37)+Success+DockQ+(1-RMSD/10))/6.

K Ablation Analysis

To assess the contribution of each component in PepMALD, seven ablation variants are evaluated. In w/o MHAE, MHAE is replaced with a basic VAE. In w/o Agg, multilayer hidden-state aggregation is removed, and the latent variables are constructed using only the final encoder layer. In w/o Global, complex-level global context injection is removed. For latent diffusion, Manifold-Anchored Latent

Diffusion is replaced with a basic LDM in w/o MALD. The LRA and IGAD modules are removed in w/o LRA and w/o IGAD, respectively, with the latter reducing the diffusion process to a uniform schedule. Finally, latent space statistical normalization is removed in w/o SN.

The results are shown in Figure 13. PepMALD achieves the highest Avg, indicating that its components jointly improve the balance among generation diversity, binding stability, and structural accuracy. Among the MHAE-related variants, the largest performance decrease is observed for w/o Agg. This result shows that constructing latent variables from only the final hidden state limits their ability to capture structural and sequence information across multiple scales. The overall performance is also reduced by w/o Global, together with declines in both Div and Con, indicating that complex-level global context injection contributes to the quality of the generated distribution and conformational consistency. Although w/o MHAE performs better on several distribution-related metrics, it remains inferior to PepMALD overall. A basic VAE may therefore produce greater apparent diversity while providing a less effective latent representation of binding structures and functional constraints. The ablations of the latent diffusion components further show that the improvement of PepMALD cannot be attributed solely to the autoencoder. Although w/o MALD achieves a Success close to that of PepMALD, it performs worse in ΔG, DockQ, and RMSD, making the generated binding conformations less reliable overall. Removing LRA also leads to a worse ΔG, indicating that the fixed latent reference helps guide the denoising trajectory toward the valid latent space. The largest decrease in Success is observed for w/o IGAD, showing that interface-guided adaptive diffusion is important for generating viable binding conformations. Although w/o SN performs better on a few individual metrics, its binding stability is substantially reduced.

Several ablation variants show advantages on individual metrics, but these improvements are generally accompanied by declines in other important measures. The overall performance of PepMALD therefore arises from the coordinated contributions of its components rather than from optimizing any single metric in isolation.

L SENSITIVITY OF PEPMALD

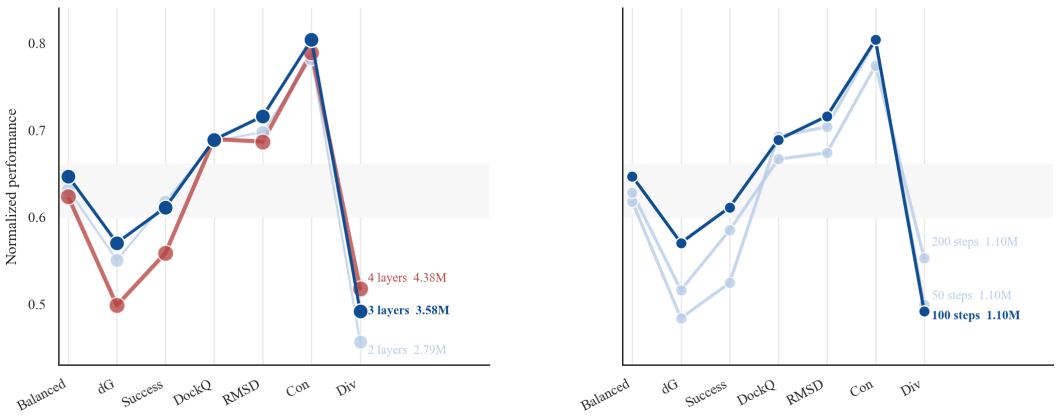

Figure 14: Sensitivity analysis of PepMALD. Normalized performance is computed after orienting all metrics so that higher values are better, using -ΔG/37 for binding energy and 1-RMSD/10 for structural deviation; the balanced score is the mean of Div, Con, -ΔG/37, Success, DockQ and 1 - RMSD/10. Marker and line sizes indicate parameter count.

To evaluate the sensitivity of PepMALD to key hyperparameters, we examine how the number of MHAE encoder layers and latent diffusion sampling steps affects performance and parameter count. As shown in Figure 14, the evaluation metrics are normalized to a common scale, and a balanced score is calculated to summarize the overall trade-off among the different criteria. Increasing the number of MHAE encoder layers from two to three clearly improves the balanced score, indicating that more complete hierarchical feature aggregation helps construct a stable latent manifold. Increasing the encoder depth to four layers, however, leads to a higher parameter count with only limited

additional improvement. Simply increasing the encoder depth therefore does not continuously improve generation quality. A similar saturation trend is observed for latent diffusion sampling steps. Using 100 steps provides better overall performance than using 50 steps, indicating that the reverse process requires sufficient sampling steps to complete manifold regression and interface residue stabilization. Increasing the sampling length to 200 steps provides no further improvement, suggesting that a longer sampling chain offers little additional benefit in incorporating receptor-conditioned information.

M Robustness across Different Peptide Characteristics

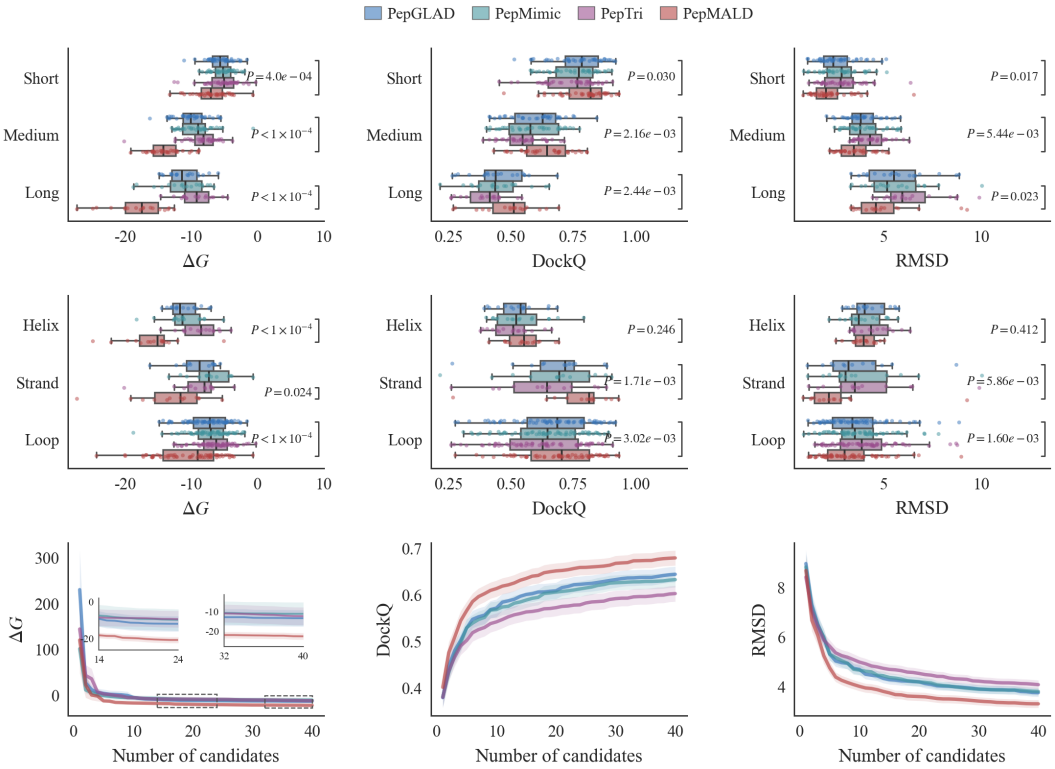

Figure 15: Robustness analysis of PepMALD across peptide secondary structures, peptide lengths, and different numbers of sampled candidates.

Peptide design samples vary substantially in structural constraints, conformational freedom, and dependence on the binding interface. Overall average performance alone is therefore insufficient to fully evaluate model robustness. Based on the DSSP-derived dominant secondary structure, the test samples are divided into Helix, Strand, and Loop groups. They are further divided by peptide length into Short, Medium, and Long groups, corresponding to 4-8, 9-15, and 16-25 amino acid residues, respectively. Within each subgroup, PepMALD is compared with the latent diffusion baselines in terms of ΔG, DockQ, and RMSD. Statistical significance is assessed by comparing PepMALD with the strongest baseline in each subgroup. A best-of-K analysis is also conducted to evaluate the ability of each model to generate high-quality designs as the number of candidates K increases.

The results are shown in Figure 15. Across the secondary-structure groups, better binding stability is consistently achieved by PepMALD, indicating that its advantage is not limited to a specific structural class. In the Strand and Loop groups, PepMALD also achieves better DockQ and RMSD results, suggesting that plausible binding modes are recovered more reliably under stronger interface dependence and greater conformational flexibility. By contrast, the stronger local conformational constraints in the Helix group lead to smaller structural differences among the models. The

improvement of PepMALD in this group is therefore observed mainly in binding stability. Across the length-based groups, a consistent advantage is maintained by PepMALD, with a particularly clear improvement in the Long group. Long peptides have a larger conformational search space and are more prone to structural drift and interface displacement during generation. Nevertheless, favorable binding stability, interface quality, and structural accuracy are preserved by PepMALD in this setting, indicating a stronger ability to model complex conformational spaces. The best-of-K analysis further shows that PepMALD consistently outperforms the baselines across all metrics as K increases. This advantage is therefore not driven by a small number of favorable samples, but is reflected in the overall quality of the candidate pool.

N Performance under Flexible Peptide Design Scenarios

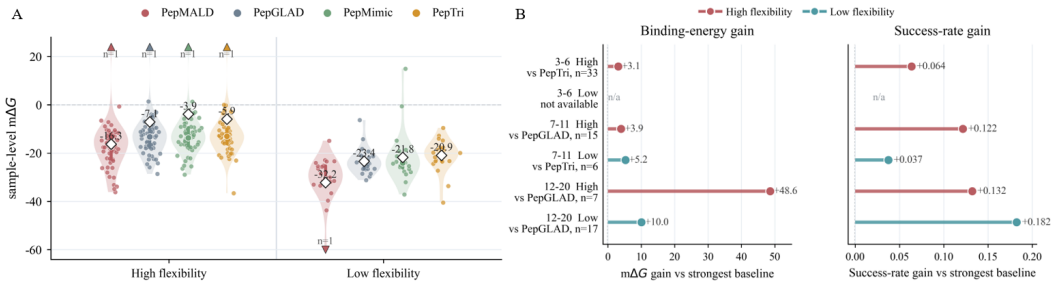

Figure 16: **A.** Sample-level $m\Delta G$ distributions of PepMALD and three baseline models on high- and low-flexibility peptides. Diamonds denote mean values. **B.** Length-controlled comparison between PepMALD and the strongest baseline in each flexibility-length subset.

Peptide flexibility substantially affects the difficulty of receptor-conditioned design. We therefore evaluate PepMALD across different flexibility regimes to determine whether its performance advantage extends to peptides with varying levels of conformational freedom.

The coil fraction in the peptide secondary structure is used as the measure of peptide flexibility. Secondary structures are predicted using PSSpred in I-TASSER (Yang & Zhang, 2015). Because coil regions lack regular hydrogen-bonding patterns, they generally have greater conformational freedom. The coil fraction is calculated for each peptide in PepBench. To compare two clearly separated flexibility regimes, peptides with a coil fraction of 1 are assigned to the high-flexibility group, whereas those with a coil fraction no greater than 0.5 are assigned to the low-flexibility group. Peptides with a coil fraction between 0.5 and 1 are excluded from this extreme-group analysis. The sample-level $m\Delta G$ distributions of PepMALD and the latent diffusion baselines are compared in both groups. To account for differences in peptide-length distribution, PepMALD is further compared with the strongest baseline within different length ranges.

As shown in Figure 16A, PepMALD achieves a lower mean $m\Delta G$ for both high-flexibility and low-flexibility peptides, indicating that its improvement in binding energy is not limited to a specific flexibility regime. For high-flexibility samples, a favorable energy distribution is maintained despite the larger conformational search space. For low-flexibility samples, the more stable structural constraints allow candidates with improved binding energies to be generated. These results show that the advantage of PepMALD extends beyond peptides with more regular structures to flexible peptides with greater conformational uncertainty. The length-controlled results further support this conclusion. As shown in Figure 16B, PepMALD generally achieves a better $m\Delta G$ than the strongest baseline across comparable flexibility-length subsets and also obtains a higher success rate. The improvement is particularly clear for longer low-flexibility peptides, indicating that stable binding optimization is maintained as the conformational search space expands with peptide length. Although the number of high-flexibility samples is smaller, PepMALD still shows an overall improvement over the baseline, supporting its robustness in flexible conformation generation.

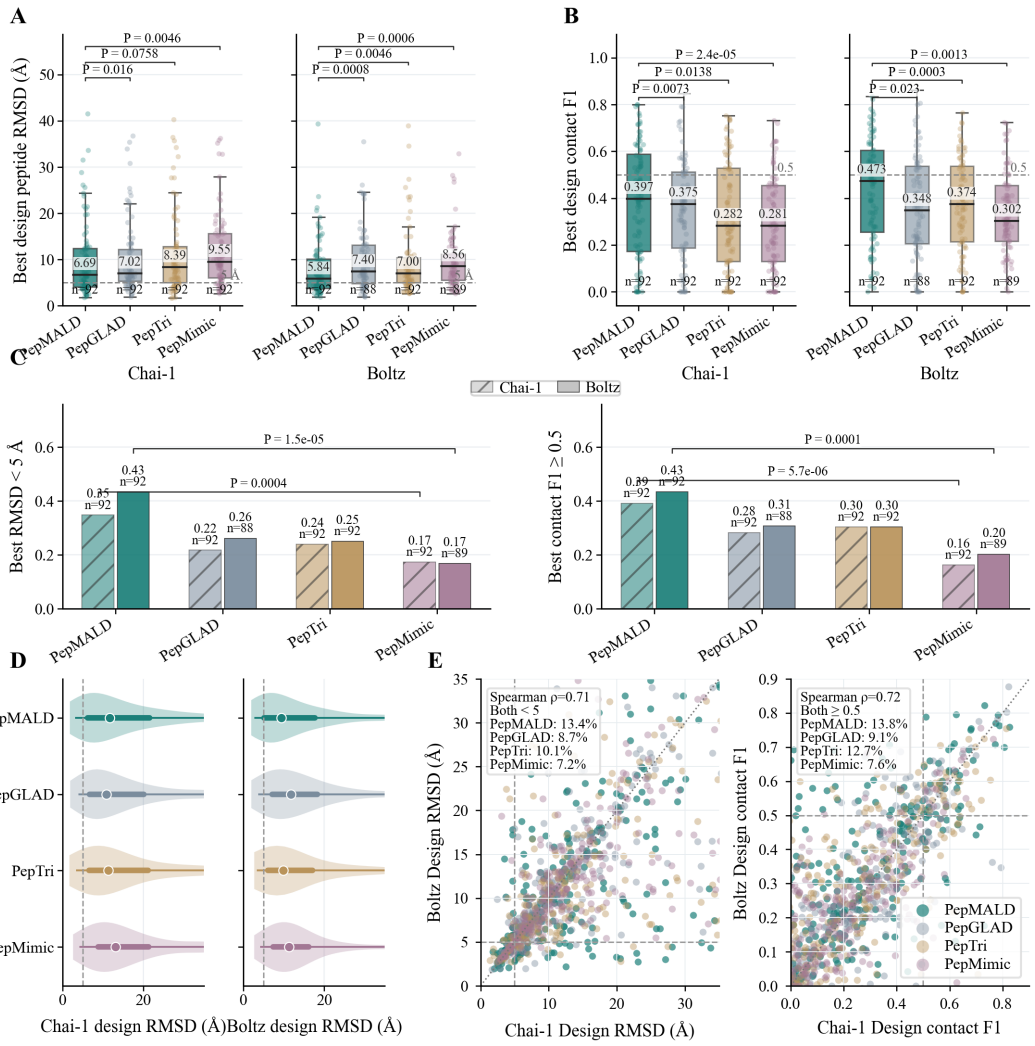

Figure 17: **A.** Target-level design-mode RMSD recovery under independent predictors. Box plots show the target-level best C_α RMSD among top-3 candidates after receptor alignment, as evaluated by Chai-1 and Boltz. **B.** Target-level interface contact recovery under independent predictors. Box plots compare the target-level best design-mode contact F1 among top-3 candidates for each method. **C.** Strict design recovery success rates. Grouped bars report the fraction of targets whose top-3 candidates reach RMSD < 5 Å or contact F1 ≥ 0.5 under Chai-1 and Boltz. **D.** Distribution of design-mode RMSD across all top-3 candidates. **E.** Concordance between Chai-1 and Boltz in design recoverability assessment. Scatter plots compare Chai-1 and Boltz predictions for the same generated designs in terms of design-mode RMSD and contact F1.

O Independent Predictor Validation of Design Recoverability

To determine whether the conformations generated by PepMALD are favored only by its internal evaluation framework, we use two independent structure predictors, Chai-1 (team et al.) and Boltz (Wohlwend et al., 2025), for external cross-validation of the generated designs. The main question is not whether these predictors can fully reproduce the native peptide conformations, but whether they can recover the binding modes proposed by the generative model using only the receptor sequence and designed peptide sequence. For each prediction, receptor alignment is performed between the predicted and designed complexes. We then calculate the design-mode RMSD, defined as the C_α RMSD of the predicted peptide relative to the designed peptide, together with the con-

tact F1. Because this protocol is independent of the PepMALD generation and screening pipeline, it provides a direct assessment of design consistency between the generated conformation and its sequence constraints.

For the 92 targets in LNR, the top-3 candidates per target are selected from PepMALD, PepGLAD, PepTri, and PepMimic, resulting in 1,104 predictor inputs. The 6seo target was excluded because its receptor contains a non-canonical modified lysine (MYK), which cannot be represented by the standard amino-acid sequence inputs required by Chai-1 and Boltz. Chai-1 successfully predicts all candidates. Boltz completes all predictions for PepMALD and PepTri, while a small number of predictions fail for PepGLAD and PepMimic.

We first compare the target-level best candidate from each model. As shown in Figure 17A, PepMALD exhibits more reliable external recoverability in terms of design-mode RMSD. Under Chai-1, the median target-level best design-mode RMSD of PepMALD is 6.69 Å, lower than the 7.02 Å obtained by the second-best method, PepGLAD. The difference is larger under Boltz, where PepMALD achieves a median of 5.84 Å, compared with 7.00 Å for the second-best method, PepTri. The paired target-level tests further support this result. Under Boltz, PepMALD achieves significantly lower design-mode RMSD than all baselines, indicating that the improvement is not driven by a small number of outlier targets.

The interface-contact recovery results provide further evidence of the design consistency of PepMALD. As shown in Figure 17B, PepMALD achieves a median target-level best contact F1 of 0.397 under Chai-1, exceeding the 0.375 obtained by the second-best method, PepGLAD. Under Boltz, its median contact F1 reaches 0.473, compared with 0.374 for the second-best method, PepTri. Paired target-level tests show that PepMALD achieves significantly higher design-mode contact F1 than all three baselines. Its external recoverability is therefore evident in the peptide backbone position and the preservation of receptor-peptide interface contacts.

PepMALD retains its advantage under stricter success thresholds. As shown in Figure 17C, based on the target-level best result among the top-3 candidates, PepMALD achieves a design-mode RMSD $<$ 5 Å for 35% of the targets and a contact F1 ≥ 0.5 for 39% of the targets under Chai-1. Under Boltz, the corresponding rates increase to 43%. The three baseline models have lower overall success rates under both threshold-based criteria.

To verify that this advantage does not arise solely from selecting the target-level best candidate, we further examine the design-mode RMSD distribution over all top-3 candidates. As shown in Figure 17D, the distribution of PepMALD is shifted toward lower RMSD values relative to PepMimic and PepTri. Under Boltz, PepMALD also achieves a lower median than PepGLAD. This distribution-level result indicates that the external recoverability of PepMALD is not limited to a small number of high-quality candidates, but reflects more stable structural consistency across the candidate set.

Finally, we assess whether Chai-1 and Boltz provide consistent evaluations of the same designs. As shown in Figure 17E, the two predictors exhibit moderate-to-high correlations for both design-mode RMSD and contact F1, with Spearman correlation coefficients of approximately 0.714 and 0.720, respectively. Although Chai-1 and Boltz are based on different model assumptions and confidence systems, they produce similar rankings of design recoverability. The advantage of PepMALD is therefore observed under both external structure predictors rather than being specific to the preference of a single predictor.

P Physics-Based Validation of Generated Binding Conformations

To further assess the physical feasibility of the complexes generated by PepMALD, two independent physical validation protocols are introduced in addition to structure-generation evaluation: Rosetta relax and OpenMM restrained molecular dynamics. Rather than relying on model-internal scores, we examine whether the generated conformations retain their original binding modes after energy optimization, force-field minimization, and short restrained simulations. Specifically, Rosetta relax is first applied to the top-3 candidates generated by each model for every LNR target. The resulting structures are evaluated using post-relax interface ΔG, peptide RMSD, and interface contact retention. Among the candidates that retain a negative ΔG and show no obvious clash after Rosetta

relax, the top-1 structure is selected for each target and subjected to OpenMM energy minimization, 50k-step restrained MD, and 100k-step restrained MD.

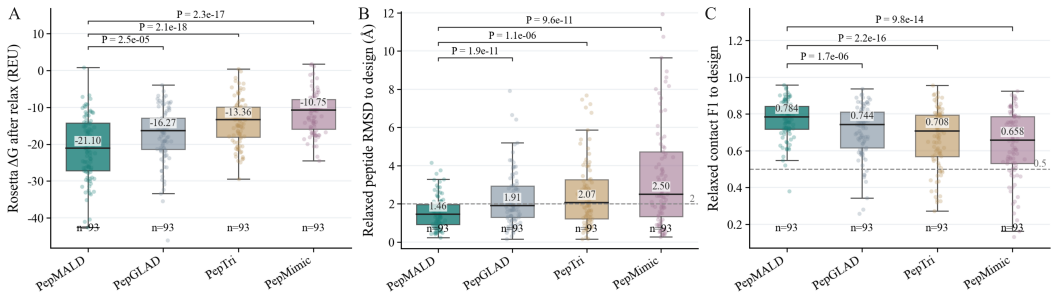

Figure 18: **A.** Rosetta-relaxed binding energy of generated complexes. **B.** Peptide conformational deviation after Rosetta relax. **C.** Interface contact preservation after Rosetta relax.

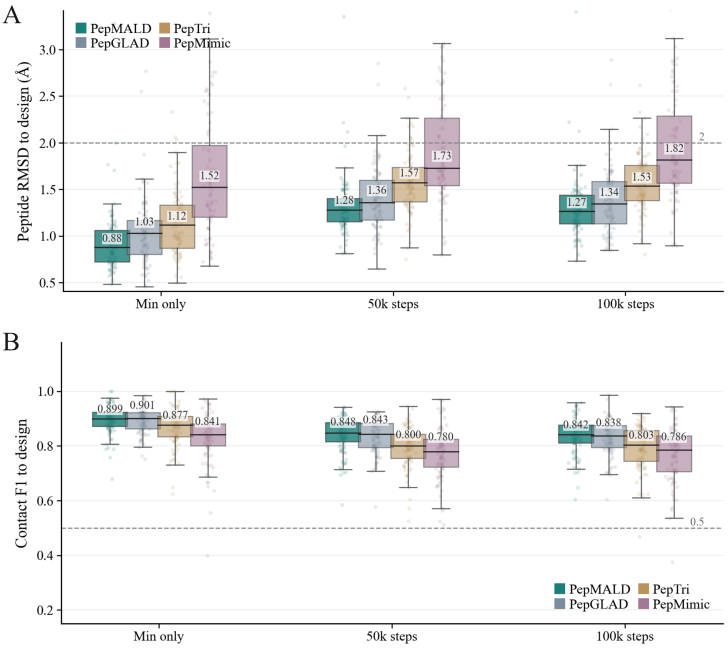

Figure 19: **A.** Peptide RMSD across OpenMM minimization and restrained MD stages. Box plots summarize peptide RMSD after energy minimization, 50k-step restrained MD, and 100k-step restrained MD. **B.** Interface contact preservation across OpenMM stages. Box plots report design-mode contact F1 after OpenMM minimization, 50k-step restrained MD, and 100k-step restrained MD.

The Rosetta relax results show that PepMALD generates candidates with more favorable interface energies. As shown in Figure 18A, when the best candidate among the top-3 candidates is considered for each target, PepMALD achieves a median post-relax interface ΔG of -21.10 REU, compared with -16.27 REU for the second-best method, PepGLAD. Paired target-level tests further show that PepMALD achieves significantly lower post-relax interface ΔG than all three baselines across the evaluated targets.

The designed conformations produced by PepMALD also undergo smaller structural changes during Rosetta relax. As shown in Figure 18B-C, PepMALD achieves a median post-relax peptide RMSD of 1.46 Å, lower than the 1.91 Å obtained by the second-best method, PepGLAD. Its median post-relax contact F1 is 0.784, exceeding the 0.744 achieved by PepGLAD. These results indicate

that the original peptide conformations and interface contacts are better preserved during physical optimization.

Under the applied OpenMM restraint protocol, PepMALD exhibits limited conformational deviation across the min-only, 50k-step, and 100k-step stages. As shown in Figure 19, its median peptide RMSD increases from 0.88 Å after minimization to 1.28 Å after 50k-step restrained MD and remains nearly unchanged at 1.27 Å after 100k-step restrained MD. The corresponding contact F1 values are 0.899, 0.848, and 0.842. Thus, extending the restrained simulation from 50k to 100k steps produces little additional structural drift or interface-contact loss under the same restraint conditions.

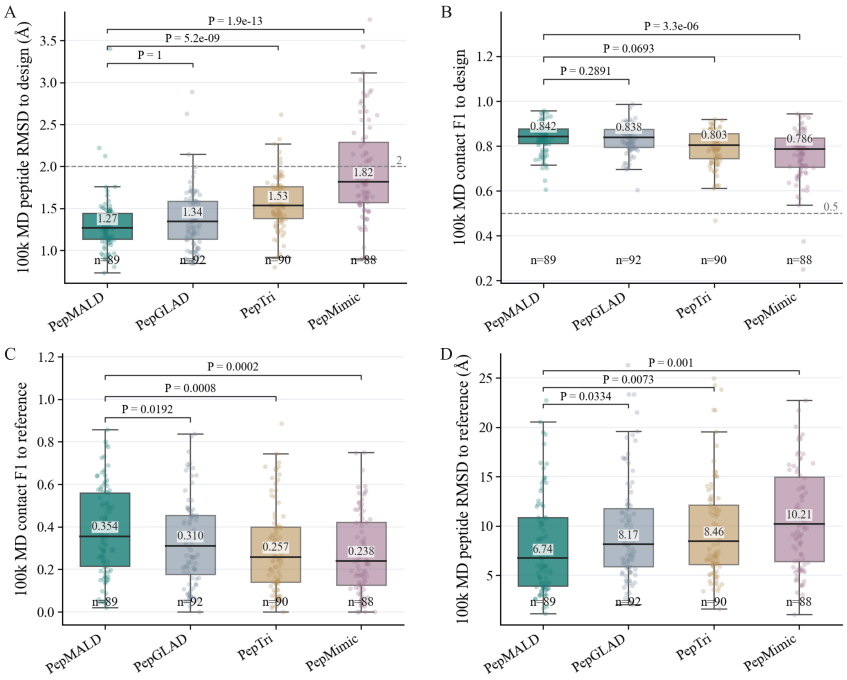

Figure 20: **A.** Final design-mode peptide RMSD after 100k-step OpenMM MD. Box plots compare peptide RMSD between the 100k-step final frame and the original design conformation. **B.** Final design-mode contact F1 after 100k-step OpenMM MD. Box plots compare interface contact F1 between the 100k-step final frame and the original design interface. **C.** Final reference-mode peptide RMSD after 100k-step OpenMM MD. Box plots report peptide RMSD between the 100k-step final frame and the reference peptide conformation. **D.** Final reference-mode contact F1 after 100k-step OpenMM MD. Box plots compare receptor-peptide contact F1 between the 100k-step final frame and the reference interface.

The final frames provide further evidence of conformational retention under the applied restraints. As shown in Figure 20, PepMALD achieves a median design peptide RMSD of 1.27 Å, lower than the 1.34 Å obtained by PepGLAD. Its median design contact F1 is 0.842, exceeding those of PepTri and PepMimic and remaining comparable to that of PepGLAD. Relative to the reference complexes, PepMALD also achieves the lowest median peptide RMSD of 6.74 Å and the highest median contact F1 of 0.354 among the compared methods. These results demonstrate comparatively stronger structural retention under the restrained-MD protocol rather than unconstrained binding stability.

Finally, quality control after 100k-step restrained MD shows that no obvious interface clash develops in any successfully completed simulation. As shown in Figure 21, 97% of the successfully simulated PepMALD samples maintain a peptide RMSD < 2 Å. All successfully simulated samples also retain a contact F1 ≥ 0.5 and remain clash-free.

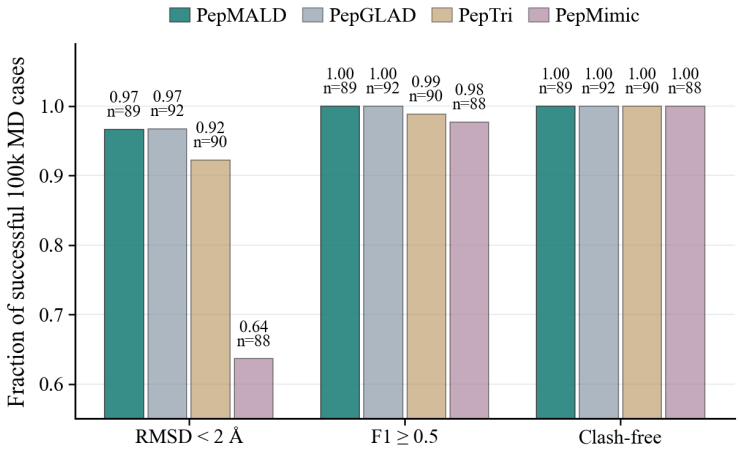

Figure 21: Quality-control success rates after 100k-step OpenMM MD. Grouped bars report the fractions of successful final-frame structures satisfying peptide RMSD < 2 Å, contact F1 ≥ 0.5, and clash-free criteria.

Q Discussion

Several limitations should be considered when interpreting the practical implications of these results. Available protein-peptide complex datasets remain limited in both scale and structural diversity. Although performance on the LNR benchmark and independent test sets indicates that PepMALD can generalize beyond the receptor clusters represented in training data, its performance on poorly characterized or underrepresented receptor families remains unclear. PepMALD also conditions generation on a single static receptor structure and uses receptor proximity to guide residue-level diffusion. This design does not explicitly capture binding-induced receptor conformational changes or distinguish spatial proximity from energetic contributions, which may limit its applicability to highly flexible interfaces. In addition, the current validation is entirely computational. While structure prediction, molecular-mechanics evaluation, and the colorectal cancer case study provide evidence for structural plausibility and binding feasibility, they cannot determine experimental performance.

Future work could improve PepMALD by expanding training coverage through newly released experimental structures and carefully selected predicted complexes. Incorporating receptor ensembles or molecular dynamics trajectories may provide a more complete description of binding-site flexibility, while interaction-aware constraints could further improve the current distance-based diffusion strategy. Ultimately, experimental synthesis and functional assays will be necessary to assess whether the computational improvements observed here can translate into practical peptide discovery. These data would also provide valuable feedback for further developing PepMALD toward real-world applications.