PPDIFF: POSITION-PRIORITY-AWARE MASKED DIFFUSION FOR CONTROLLABLE PEPTIDE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Controllable peptide generation requires careful refinement of residues that determine the requested properties. Masked discrete diffusion offers flexible generation trajectories, but uniform training corruption allocates reconstruction practice independently of downstream property relevance. We introduce **PPDIFF**, a controllable generation framework that uses an external property planner at both stages. A lightweight sequence random forest with in silico mutagenesis estimates residue criticality and uncertainty. The resulting prior reallocates a fixed training mask budget and, during sampling, is combined with denoiser confidence, margin, and temporal stability to decide when each residue is ready to commit. The selected pretrained masked-diffusion backbone is retained without architectural changes, and models used for evaluation are separate from the guide. Across permeability, hemolysis safety, and nonfouling, **PPDIFF** improves the joint objective under three separate sequence evaluators while supporting target-specific candidate design in computational screens against seven proteins. A multi-seed factorial study and a shuffled-prior control further show that the main gain arises when property-bearing position information coordinates both training corruption and sampling commitment.

1 INTRODUCTION

Peptides can recognize extended protein surfaces while remaining comparatively accessible to synthesis and sequence-level design (Muttenthaler et al., 2021). Designing useful peptides requires controlling coupled attributes such as membrane permeability, hemolytic safety, surface nonfouling, and target binding. Residue contributions are position dependent: in aurein 1.2, alanine substitutions at two neighboring lysines have opposite effects on hemolysis (Migoń et al., 2019). A controllable generator should therefore devote particular care to residues that determine the requested properties as it constructs a sequence. The challenge is to translate this unequal residue importance into decisions about where to concentrate generation effort.

Masked discrete diffusion models (MDMs) expose a mechanism for addressing this challenge. At each denoising step, an MDM predicts residues from a partially masked sequence, and its sampler selects which positions to reveal next (Austin et al., 2021; Sahoo et al., 2024; Shi et al., 2024). This flexible order allows commitment at uncertain positions to be deferred as sequence context accumulates, making where and when to resolve residues natural points of control for peptide properties. DPLM (Wang et al., 2024) establishes this approach for protein sequences; adapting it to peptide design calls for linking its position choices to peptide-specific objectives. Yet standard MDM training masks positions uniformly at a given noise level, independently of those objectives.

Making the commitment path property dependent only at inference leaves reconstruction practice indifferent to the positions on which property-directed generation relies most. This creates a mismatch in how modeling effort is allocated across the two stages. Path Planning (Peng et al., 2025) demonstrates the value of choosing denoising orders for biological sequences, and PAPL (Peng et al., 2026) formalizes a related mismatch between uniform-path training and planned inference. Information-density masking (Ma et al., 2026) further supports selective training corruption. For controllable peptide generation, this yields two coupled requirements: identifying positions important to the target property and using that information consistently across training and sampling.

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

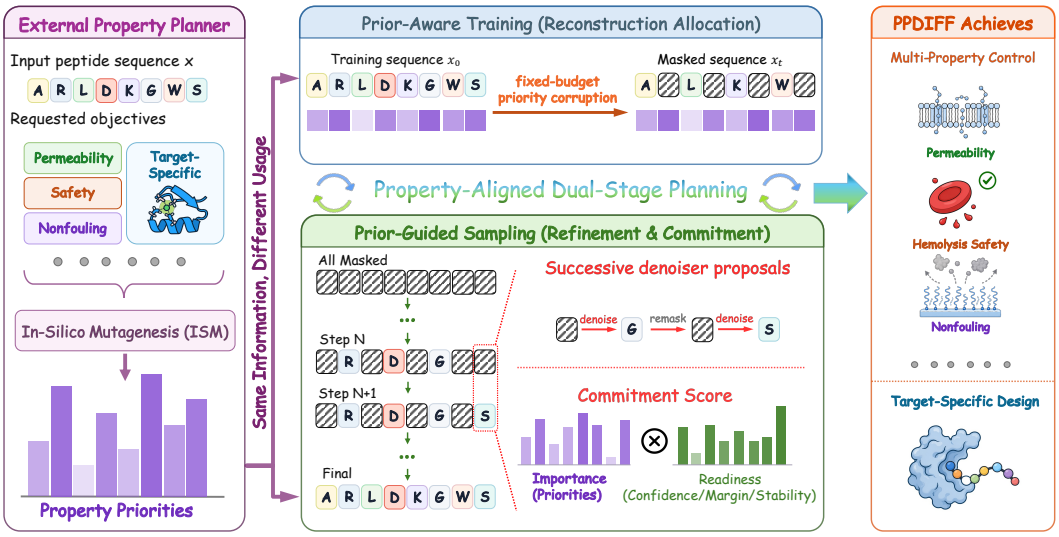

Figure 1: **Overview of PPDIFF: property-aligned dual-stage planning. Left:** An external scorer coupled with *in silico* mutagenesis derives residue priorities, optionally conditioned on a target protein sequence. **Top:** Prior-aware training reallocates a fixed corruption budget toward high-priority positions. **Bottom:** Joint-readiness sampling balances priorities with token readiness (confidence, margin, and stability) to iteratively steer refinement across remasking and denoising cycles. **Right:** Capability across diverse tasks, including multi-property control and target-specific design.

We introduce **PPDIFF**, a framework that addresses both requirements with an external property planner and a shared position prior (Figure 1). A lightweight sequence random forest (Breiman, 2001) scores each property, and in silico mutagenesis (ISM) (Nair et al., 2022) measures the effects of residue substitutions. Combining these effects with predictor uncertainty yields a property-informed position prior. During fine-tuning, this prior reallocates a fixed mask budget toward important positions, concentrating reconstruction practice on them. During sampling, the same prior is combined with denoiser confidence, top-one margin, and temporal stability to guide commitment while allowing unstable proposals to undergo further refinement. Thus, a shared property objective coordinates reconstruction practice and residue refinement; it does not require the two stages to visit identical partial states.

Across permeability, hemolysis safety, and nonfouling, **PPDIFF** improves joint property control with consistent method rankings under separate protein-language-model, fingerprint, and Peptide-BERT (Guntuboina et al., 2023) evaluators. A seven-target computational study extends this evaluation to target-specific peptide design. A multi-seed factorial experiment finds a positive interaction when an informative position prior controls both stages, whereas position-shuffled priorities yield nearly additive gains. This contrast supports the value of coordinating training and sampling through property-relevant position information.

Our contributions are twofold. **First**, we develop an external planner that translates property scores into residue priorities and accommodates both property-specific and target-conditioned scorers. **Second**, we introduce shared-prior control of fixed-budget training corruption and sampling commitment, and establish its benefit through a multi-seed 2×2 design, shuffled-prior controls, and commitment ablations.

2 Related Work

Masked diffusion for sequences and proteins. Multinomial diffusion (Hoogeboom et al., 2021) and D3PM (Austin et al., 2021) extend denoising to categorical state spaces, with the latter explicitly supporting absorbing mask transitions. MDLM (Sahoo et al., 2024) reduces masked diffusion training to weighted masked-language-model losses, and generalized masked diffusion (Shi et al., 2024) permits state-dependent masking schedules. LLaDA (Nie et al., 2025) demonstrates scaling to general-

purpose language modeling. In biology, EvoDiff (Alamdari et al., 2023) generates proteins through sequence-space discrete diffusion, DPLM (Wang et al., 2024) pretrains diffusion protein language models at scale, and MemDLM (Goel et al., 2024) specializes masked diffusion and token-level guidance for membrane proteins. These models provide sequence backbones on which **PPDIFF** builds property-directed trajectories for peptide generation.

Non-uniform paths and planned inference. Path Planning (Peng et al., 2025) studies how token order affects MDM sampling, including protein and RNA generation. PAPL (Peng et al., 2026) shows that planned inference is not faithfully described by the uniform-path ELBO and derives a planner-aware objective. Information-density masking (Ma et al., 2026) preferentially corrupts informative spans during diffusion-language-model training. **PPDIFF** brings these trajectory considerations to controllable peptide generation: an external planner estimates property relevance at each position, and the resulting prior controls both training corruption and sampling commitment.

Controllable peptide generation and evaluation. Classifier guidance (Gruver et al., 2023; Nisonoff et al., 2025) steers discrete sequence generators using property predictors, while discrete denoising posterior prediction (DDPP) (Rector-Brooks et al., 2025) learns to sample from reward-tilted distributions of a pretrained MDM. HydrAMP (Szymczak et al., 2023) uses a conditional variational autoencoder for antimicrobial peptide generation. For peptide design, PepMLM (Chen et al., 2026) generates binders conditioned on target sequences and scores candidates by leave-one-residue-out pseudo-perplexity. PepTune (Tang et al., 2025) combines peptide-SMILES diffusion with tree search over therapeutic objectives. **PPDIFF** uses property scores to derive position priorities that coordinate reconstruction during training and residue commitment during sampling.

3 METHOD

Figure 1 summarizes the resulting pipeline and the distinct roles of the shared position prior.

3.1 PROBLEM SETTING AND MASKED-DIFFUSION BACKBONE

Let $x_0 = (x_0^1, \ldots, x_0^L) \in \mathcal{V}^L$ be a peptide over the 20 canonical amino acids, and let p denote a requested property. We seek samples with high property scores while preserving the sequence quality learned by a pretrained MDM. Our implementation uses DPLM-150M as a representative protein-sequence backbone. **PPDIFF** adds a position prior, a training corruption rule, and a sampling commitment rule while retaining the backbone and token-head architecture.

For an absorbing mask state $\mathbf{m}$ and noise level α_t, the standard property-independent corruption marginal is

$$q(x_t^i \mid x_0^i) = (1 - \alpha_t)\,\delta(x_t^i = x_0^i) + \alpha_t\,\delta(x_t^i = \mathbf{m}). \tag{1}$$

The denoiser $p_\theta(x_0 \mid x_t, t)$ predicts masked residues (Sahoo et al., 2024; Wang et al., 2024). Standard training chooses corrupted coordinates independently of property relevance. **PPDIFF** instead gives training corruption and sampling commitment access to the same external property planner.

3.2 PROPERTY-CONDITIONED POSITION PRIOR

For each sequence property p, we train a sequence random forest f_p on padded one-hot peptides and exclude padding from attribution. Appendix A.1 specifies the guide data and splits; separate models evaluate generated sequences. Given sequence x, ISM substitutes residue x_i with every alternative amino acid a and computes

$$\Delta_{i,a}^{(p)} = f_p(x^{i\to a}) - f_p(x), \qquad \bar{\Delta}_i = \frac{1}{19}\sum_{a \neq x_i} \Delta_{i,a}^{(p)}, \qquad c_i = -\bar{\Delta}_i. \tag{2}$$

Positive c_i indicates that substitutions tend to reduce the predicted property. Cross-tree disagreement in criticality, min–max normalized within each sequence, gives uncertainty $u_i \in [0, 1]$ (Appendix B.1). The priority is

$$q_i^{(p)}(x) = c_i^{(p)}(x) - \lambda_u u_i^{(p)}(x). \tag{3}$$

The priority ranks positions by guide-score sensitivity after uncertainty discounting.

For multiple properties, every guide is oriented so that larger values are preferred. Hemolysis is therefore represented by safety. We percentile-normalize each priority within the current sequence and combine requested properties as

$$q_i^{(\mathbf{w})}(x) = \sum_{p \in \mathcal{P}} w_p \left[\mathrm{PctRank}\Big(q^{(p)}(x) \Big) \right]_i, \quad w_p \geq 0, \quad \sum_p w_p = 1. \tag{4}$$

The experiments use Equations 3 and 4 and assign equal weights to all target properties by default.

3.3 Target-conditioned binding prior

The same interface also accepts a scorer conditioned on auxiliary context. For a target protein sequence z and candidate peptide x, we use the frozen `PepMLM-650M` (Chen et al., 2026). Following PepMLM's public scoring procedure, each peptide residue is masked once while the target and all remaining peptide residues are observed:

$$\mathrm{PPL}_\phi(x \mid z) = \exp\left[-\frac{1}{L} \sum_{i=1}^{L} \log p_\phi\big(x_i \mid z \,\|\, x_{\setminus i}\big) \right]. \tag{5}$$

Lower pseudo-perplexity indicates target-conditioned sequence compatibility. To obtain a common rank scale, we fix $\mathcal{R}_z$ before generation as the same validation-split canonical panel for every target, filtered to the study's length support. We define

$$b_\phi(z, x) = 1 - \widehat{F}_{\mathcal{R}_z}\left(\log \mathrm{PPL}_\phi(x \mid z)\right), \tag{6}$$

where $\widehat{F}_{\mathcal{R}_z}$ is the empirical CDF of target-specific reference pseudo-perplexities. We substitute $b_\phi(z, x)$ for $f_p(x)$ in Equation 2, set $u_i^{(B)} = 0$ for this deterministic guide, and combine its percentile-ranked mutation priority with the three property priorities using equal weights. Each target receives a separate fine-tuning-and-sampling trajectory from the same checkpoint because z conditions the position prior but is not supplied to the backbone token predictor. AutoDock Vina is reserved for subsequent screening of frozen candidate pools. Exact ISM with this score requires $L(19L + 1)$ masked-input evaluations per refresh before caching; this cost is separate from RF guidance (Appendix B.6.1).

3.4 Fixed-budget priority corruption

As illustrated in the upper center of Figure 1, at time t we retain the standard mask count $M_t = \lceil \alpha_t L \rceil$ and change only the selected coordinates. Let C_i mark the top priority fraction f. During epoch e, the curriculum coefficient ρ_e interpolates between uniform and property-aware weights:

$$C_i = \mathbb{I}[q_i \geq \mathrm{quantile}(q, 1 - f)], \qquad \pi_i^{(e)} = (1 - \rho_e)\frac{1}{L} + \rho_e \frac{wC_i + (1 - C_i)}{\sum_j [wC_j + (1 - C_j)]}. \tag{7}$$

Stochastic policies sample exactly M_t positions without replacement using the Gumbel top-k construction (Kool et al., 2019):

$$S_t = \mathrm{TopK}_{i \in [L]} \left\{ \log \pi_i^{(e)} + G_i ; M_t \right\}, \quad G_i \sim \mathrm{Gumbel}(0, 1), \quad |S_t| = M_t. \tag{8}$$

The headline `property_ordered` policy directly selects the M_t largest continuous priorities, whereas (ρ_e, w, f) apply only to stochastic and binary ablations. In every case, the objective remains masked-token cross-entropy on the selected subset:

$$\mathcal{L}_{\mathrm{train}}(\theta) = \mathbb{E}_{t, x_0, S_t} \left[a(t) \sum_{i \in S_t} \mathrm{CE}\big(p_\theta(x_0^i \mid x_t, t), x_0^i\big) \right]. \tag{9}$$

The nonnegative time weight $a(t)$ is inherited from the backbone (Appendix B.1). Under deterministic ordering, a position with descending priority rank ϱ_i is reconstructed whenever $M_t \geq \varrho_i$, giving exposure probability $E_i^{\mathrm{train}} = \Pr(M_t \geq \varrho_i)$. High-priority residues therefore receive reconstruction practice across more noise levels while the mask count at each level stays fixed. Tables 7 and 12 report the policy variants and default configuration.

3.5 IMPORTANCE AND READINESS DURING SAMPLING

Sampling combines property importance with current-proposal reliability (Figure 1, bottom). At sampling step t, let $p_i^{(t)}$ be the denoiser distribution at a mutable residue position. We measure its confidence, top-one margin, and temporal stability as

$$\mathrm{conf}_i = \max_a p_i^{(t)}(a), \quad \mathrm{margin}_i = p_i^{(t),(1)} - p_i^{(t),(2)}, \quad \mathrm{stab}_i = 1 - \mathrm{JSD}_2\Big(p_i^{(t)}, p_i^{(t-1)}\Big). \quad (10)$$

Here JSD_2 uses base-2 logarithms, and stability is one at the first step. The prior is recomputed on the current full proposal sequence $\hat{x}_t$ every M steps. With $\sigma(z) = (1 + e^{-z})^{-1}$, the importance gate, readiness, and commitment score are

$$g_i(t) = \sigma(q_i(\hat{x}_t)/\tau_g), \quad r_i(t) = \mathrm{conf}_i^{\alpha}\,\mathrm{margin}_i^{\beta}\,\mathrm{stab}_i^{\gamma}, \quad s_i(t) = g_i(t)^{\lambda_g} r_i(t), \qquad \lambda_g \geq 0. \quad (11)$$

The base reparameterized sampler ranks positions by $s_i(t)$ and uses the lowest-ranked positions to fill the remaining-mask budget k_t (Appendix B.1). Importance prioritizes consequential sites, while readiness downweights uncertain proposals. The default $\lambda_g = 1$ gives the multiplicative rule, $\lambda_g = 0$ is readiness-only, and larger values increase property importance relative to readiness.

We evaluate a conservative `weighted_T_V` operator and the stronger `improvable_remask` operator used for the headline result. The latter ranks eligible positions with $h_i > 0$ by the signed mean mutation effect $h_i = \sum_p w_p \bar{\Delta}_i^{(p)}$, with a quota tied to the remaining mask budget. Selected unmasked residues return to the mask state and receive new proposals at the next denoising step; selected masked positions can remain masked. Thus commitment combines criticality with proposal readiness, while remasking uses signed mutation effects to allocate another denoising attempt. Appendix B.1 defines both operators and the complete update order. Default sampling uses $T = 64$, guide interval $M = 5$, $\tau_g = \lambda_g = 1$, and $\alpha = \beta = \gamma = 1$. Tables 12 and 19 provide the full operating point and sensitivity summary. Appendix F.2 evaluates a wider $T \times \lambda_g$ grid.

4 MAIN RESULTS

4.1 EVALUATION OVERVIEW

We evaluate permeability, hemolysis safety, and nonfouling using 1,000 raw generations per condition. Headline means use PepBenchmark ESM-150M-F, with fingerprint and PeptideBERT cross-checks. Baselines include the backbone start, continued uniform training, shuffled priorities, gradient guidance, PAPL, and PepTune. Gradient guidance steers token logits using separate differentiable property classifiers (Appendix A.3). Property means use scoreable outputs, with protocol Valid reported separately. Appendix A gives the protocol, data construction, and reproducibility details.

4.2 THE POSITION PRIOR PRODUCES CONTROLLABLE SHIFTS

Training with priority masking improves all three properties, whereas continued uniform training and position shuffling remain close to the start model (Table 1). The strongest training policy improves permeability by 0.1018, safety by 0.0743, and nonfouling by 0.0628. The consolidated distribution analysis in Table 24 finds significant shifts for the true prior and nonsignificant shifts for both controls, while Table 7 collects all training-policy and position controls. The gain therefore depends on the property-bearing coordinate ranking rather than additional fine-tuning alone.

Sampling on the matched priority-trained backbone further raises every property. `weighted_T_V` produces scores from 0.8196 to 0.8372, while `improvable_remask` produces scores from 0.8983 to 0.9102. On the property-neutral backbone, the guided sampler boosts balanced utility from 0.6273 to 0.7644. This gap motivates our factorial analysis in Section 5.1.

4.3 BALANCED MULTI-PROPERTY IMPROVEMENT

For matched evaluator scores, balanced utility is $J_o(x) = [s_{o,P}(x) + s_{o,S}(x) + s_{o,N}(x)]/3$. On this metric, **PPDIFF** reaches 0.9034 with a single-seed sequence-bootstrap 95% CI of $[0.8970, 0.9098]$. Joint satisfaction uses matched per-sequence P/S/N triples under validation-fixed thresholds (Table 2).

Table 1: Per-property control under the primary held-out evaluator. Entries are means over scoreable generations.

Condition	Perm. ↑	Safety ↑	Nonfoul. ↑
Backbone start	0.5996	0.6235	0.6587
Uniform-mask train	0.6008	0.6261	0.6603
Shuffled-prior train	0.6072	0.6297	0.6641
PPDIFF train only	0.7014	0.6978	0.7215
PPDIFF + `weighted_T_V`	0.8196	0.8213	0.8372
PPDIFF full	**0.9017**	**0.8983**	**0.9102**

Table 2: Per-sequence multi-property evaluation from matched P/S/N score triples. Joint pass is $\Pr(H_o = 1)$ under validation-fixed thresholds. The symbol ρ denotes the range of the three pairwise Spearman correlations.

Method	Joint pass ↑	Mean min. ↑	Median min. ↑	Pairwise ρ range
Backbone start	0.1543	0.5562	0.5631	0.2417–0.3586
PPDIFF training only	0.3718	0.6624	0.6712	0.3198–0.4687
PPDIFF sampling only	0.5364	0.7148	0.7226	0.3694–0.5078
PepTune	0.5581	0.7208	0.7283	0.3812–0.5189
PAPL + matched sampler	0.6758	0.7714	0.7792	0.4215–0.5578
PPDIFF full	**0.8462**	**0.8681**	**0.8757**	0.4586–0.5813

The full method reaches joint pass 0.8462 and mean worst-property score 0.8681, compared with 0.6758 and 0.7714 for PAPL+matched. Training-only and sampling-only balanced utilities are 0.7069 and 0.7644, respectively.

The requested-weight sweep shows selective property control (Appendix D.1). Across seven complete generation configurations, single-objective requests attain permeability 0.9387, safety 0.9352, or nonfouling 0.9416 for the selected property. Equal weights give the highest three-property joint pass, 0.8462, compared with 0.3624–0.3847 for single-objective settings, illustrating the trade-off between maximizing one property and satisfying all three.

4.4 Cross-representation evaluation and sequence quality

Both cross-check evaluators preserve the full method's ranking on every property (Table 3). Its balanced fingerprint score is 0.8462 and its mean PeptideBERT marginal pass fraction is 0.8494, compared with 0.7023 and 0.6837 for sampling only. Tables 30, 32, and 33 report leakage, evaluator reliability, and applicability-domain audits.

Control produces modest concentration: relative to the backbone start, Valid changes from 0.862 to 0.849, uniqueness from 0.908 to 0.898, and pairwise diversity from 0.7823 to 0.7486. Within the residue-token DPLM family, validation PPL remains comparable at 2.421 versus 2.451. Appendix D.4 gives the full quality analysis. Appendix D.1 reports objective-weight control, Appendix B.3 compares priority directions, and Appendix B.6.2 reports measured generation profiles.

4.5 Target-specific design under property constraints

Beyond property-only generation, we apply **PPDIFF** to seven targets while retaining permeability, safety, and nonfouling priorities. PepMLM pseudo-perplexity guides both target-specific training and sampling. The top-20 pools are frozen before any Vina query. Six targets show better ligand efficiency than matched decoys after false-discovery-rate correction, indicating enrichment within the computational screen; RBX1 is nonsignificant. Table 4 reports one lead per target, Figure 2 shows three pose hypotheses, and Appendix G gives the complete screen and provenance audit.

Table 3: Unified evaluation. Valid is protocol validity for residue models and canonical-protocol yield for PepTune. Property scores use scoreable outputs. ESM-150M-F and fingerprint (FP) entries are means, while PeptideBERT (PB) entries are pass fractions. P, S, and N denote permeability, safety, and nonfouling. PPL_{val} evaluates residue-model checkpoints on validation data; PepTune's SPE/SMILES generated-set PPL appears separately in Appendix A.

			ESM-150M-F			Fingerprint			PeptideBERT		
Method	Valid ↑	PPL_{val} ↓	P ↑	S ↑	N ↑	P ↑	S ↑	N ↑	P ↑	S ↑	N ↑
Backbone start	0.862	2.451	0.5996	0.6235	0.6587	0.5691	0.5724	0.5812	0.5483	0.5594	0.5517
PepTune	0.843	–	0.7700	0.7719	0.7659	0.7089	0.7176	0.7139	0.7050	0.7081	0.7065
PAPL	0.858	2.449	0.7352	0.7234	0.7410	0.6578	0.6652	0.6614	0.6497	0.6554	0.6582
PAPL + matched guided sampler	0.851	2.449	0.8190	0.8084	0.8000	0.7516	0.7584	0.7698	0.7538	0.7594	0.7713
Gradient classifier guidance	0.847	2.451	0.7118	0.7008	0.7162	0.6447	0.6386	0.6513	0.6365	0.6417	0.6489
PPDIFF training only	0.855	**2.421**	0.7014	0.6978	0.7215	0.6421	0.6468	0.6487	0.6284	0.6357	0.6328
PPDIFF sampling only	0.852	2.451	0.7498	0.7687	0.7746	0.6974	0.7039	0.7056	0.6805	0.6857	0.6849
PPDIFF full	0.849	**2.421**	**0.9017**	**0.8983**	**0.9102**	**0.8376**	**0.8518**	**0.8492**	**0.8413**	**0.8496**	**0.8574**

Table 4: Consolidated target-specific leads. Arrow symbols (↑/↓) indicate the direction of a more favorable score. Best and Cand.−ctrl are in kcal/mol; LE is in kcal/mol per heavy atom of the uncapped linear canonical peptide. Values are interpreted within each target.

Target	Sequence	Best ↓	LE ↓	Cand.−ctrl ↓	P ↑	S ↑	N ↑	Min. ↑
GFAP	`WEYLRKFWDQHKRYLF`	-10.54	-0.06311	-1.2010	0.8247	0.8672	0.8513	0.8247
TfR	`YHWRPRQFNDKLVHY`	-13.95	-0.09426	-1.0780	0.8154	0.8426	0.8347	0.8154
GLP1R	`WEFLKQYADLRKAFHE`	-14.82	-0.09946	-2.3270	0.8762	0.9043	0.8917	0.8762
NCAM1	`WDQYHRFNKELSPYQ`	-12.72	-0.08833	-0.9518	0.8429	0.8843	0.8706	0.8429
AMHR2	`KWYFRHVLQKDYFNH`	-12.47	-0.08313	-1.4380	0.8588	0.8583	0.8447	0.8447
GLAST	`MYSPVYPYEVDFTYKQFR`	-13.98	-0.08422	-0.7412	0.8491	0.8306	0.8176	0.8176
RBX1	`WYDEWWEEVKWFPYD`	-10.50	-0.06646	-0.2218	0.7518	0.7614	0.7536	0.7518

5 Mechanism Analysis

5.1 The two stages exhibit a positive interaction

We cross the absence versus presence of priority fine-tuning with baseline versus guided sampling. Each training level uses one fixed checkpoint across its two sampling cells, and each cell has five generation-seed replicates. Table 5 finds positive excess over additivity for every property. The equal-weight interaction is $+0.0592$, with a reported seed-bootstrap 95% CI of $[+0.0192, +0.0992]$. This interval reflects generation-seed variation, while independent-checkpoint variation is evaluated below.

To quantify independent-checkpoint variability, Table 14 reports three independent base/fine-tuning seeds with three generation seeds per checkpoint. The observed checkpoint-level standard deviations range from 0.0102 to 0.0237. The full-versus-PAPL+matched balanced-utility gap is 0.0931.

We next repeat the matched factorial comparison with five independent fine-tuning seeds and three generation seeds per checkpoint (Table 6). The intact-prior interaction remains positive, while the shuffled-prior interval crosses zero; their point-estimate difference is 0.0566. This contrast supports shared-priority coordination across training runs: reconstruction practice is most useful on coordinates emphasized during property-directed inference.

5.2 Policy and commitment ablations

The training-policy ablation selects `property_ordered` because it is strongest on all three properties. Continuous position priorities outperform their binarized counterparts, while mixing half

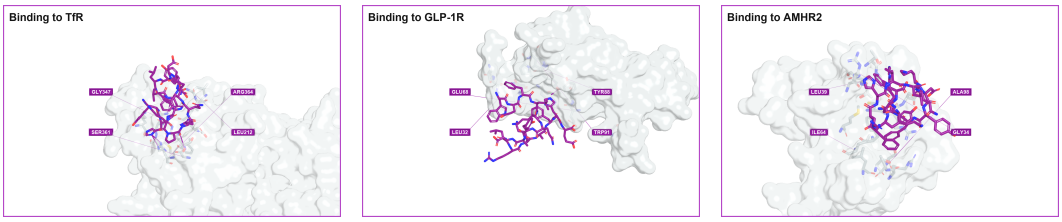

Figure 2: Pose reconstructions for illustrative leads. The receptor surface is gray, peptide atoms are magenta, and labeled residues lie within 4.2 Å of the pose.

Table 5: Factorial interaction over five generation seeds per cell, with one fixed checkpoint for each training level. Interaction equals both minus training only minus sampling only plus the backbone start. Bootstrap resampling is performed over seed summaries within cells.

Objective	Start	Train only	Sample only	Both	Interaction	95% CI
Permeability	0.5987	0.7013	0.7498	**0.9017**	+0.0493	$[+0.0095, +0.0890]$
Safety	0.6239	0.6961	0.7687	**0.8971**	+0.0562	$[+0.0178, +0.0945]$
Nonfouling	0.6583	0.7213	0.7746	**0.9097**	+0.0721	$[+0.0260, +0.1179]$
Joint (P/S/N)	0.6270	0.7062	0.7644	**0.9028**	+0.0592	$[+0.0192, +0.0992]$

Table 6: Balanced-utility factorial with five independent fine-tuning seeds and three generation seeds per checkpoint. Intervals are for each interaction; $\Delta\mathcal{I} = \mathcal{I} - \mathcal{I}_{\text{shuffled}}$ is the point-estimate difference.

Position prior	Start	Train only	Sample only	Both	$\mathcal{I}$	95% CI($\mathcal{I}$)	$\Delta\mathcal{I}$
Intact property prior	0.6274	0.7052	0.7646	0.9009	+0.0585	$[+0.0324, +0.0846]$	+0.0566
Within-sequence shuffled	0.6274	0.6348	0.7237	0.7330	+0.0019	$[-0.0142, +0.0180]$	reference

of the batches with uniform corruption weakens control. Table 7 consolidates this comparison with the random-training, shuffled-prior, and mask-complement controls.

Table 7: Consolidated training-only mask-policy ablation and position controls under ESM-150M-F. Δ is the balanced-utility difference from the backbone start. Higher is better.

Policy	Perm.	Safety	Nonfoul.	Mean	Δ
Backbone start	0.5996	0.6235	0.6587	0.6273	–
property ordered	**0.7014**	**0.6978**	**0.7215**	**0.7069**	**+0.0796**
weighted priority	0.6812	0.6786	0.7023	0.6874	+0.0601
hard top 40%	0.6693	0.6714	0.6897	0.6768	+0.0495
ordered + 50% random	0.6647	0.6653	0.6847	0.6716	+0.0443
soft priority 40%	0.6583	0.6607	0.6813	0.6668	+0.0395
random continued training	0.6008	0.6261	0.6603	0.6291	+0.0018
shuffled prior	0.6072	0.6297	0.6641	0.6337	+0.0064
mask complement	0.5991	0.6219	0.6574	0.6261	-0.0012

Importance and readiness are also complementary. On a fixed priority-trained backbone, importance alone averages 0.8365 across the three properties, additive fusion averages 0.8823, and the multiplicative rule reaches 0.9034. Removing margin causes the largest mean loss, followed by stability and uncertainty discount. Table 8 reports the full rule decomposition. Additional operator controls separate commitment from rewriting: utility is 0.8260 without remasking, 0.8573 with random remasking, and 0.9034 for the full operator (Appendix B.4). Increasing the rewrite quota beyond the default reduces utility despite more revisits. These comparisons support informative position selection and controlled refinement.

5.3 Robustness to guide and optimization shortcuts

Appendix F tests guide degradation, added sampling compute, and denoiser-free proxy optimization. The RF guides reach macro AUROC 0.8077 and mean within-sequence mutation-rank agreement

Table 8: Commit-rule ablation on a fixed priority-trained backbone under ESM-150M-F. Only the position-ranking signal changes. In the additive rule, z standardizes over mutable residue positions within each sequence at each step (Appendix B.1). P, S, and N denote permeability, safety, and nonfouling.

Commit rule	P ↑	S ↑	N ↑
Confidence only	0.7248	0.7131	0.7302
Readiness only	0.7512	0.7315	0.7466
Importance only	0.8427	0.8248	0.8420
Additive $z(g_i) + z(r_i)$	0.8879	0.8756	0.8834
Full without stability	0.8904	0.8835	0.8943
Full without margin	0.8792	0.8680	0.8832
Full without uncertainty discount	0.8948	0.8951	0.9029
Multiplicative $g_i r_i$	**0.9017**	**0.8983**	**0.9102**

0.4127 with the separate evaluator ensemble. Utility falls as local guide information is removed: within-sequence shuffling reduces multi-seed dual-end utility from 0.9009 to 0.7330. Increasing sampling-only compute reaches 0.8214 at $T = 256$, below 0.9034 for the full method at $T = 64$, with lower valid-and-joint yield. RF-only greedy search maximizes the guide more strongly but transfers poorly and raises $\mathrm{PPL}_{\mathrm{gen}}$ to 11.327. Together, these controls link performance to informative local rankings, priority-aware reconstruction practice, and denoiser proposal quality (Figure 3).

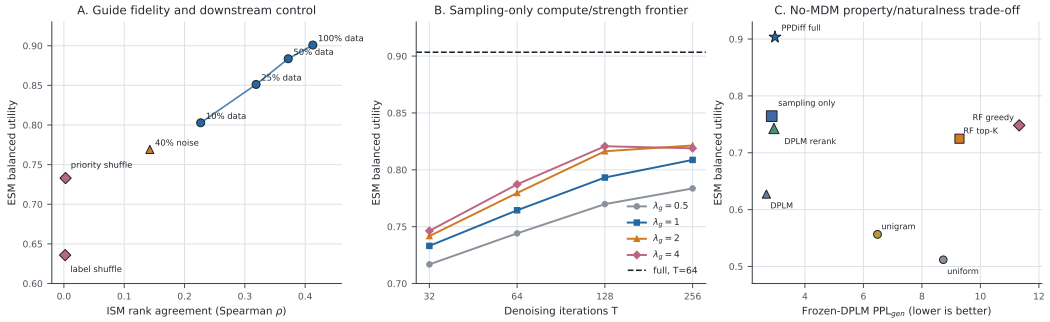

Figure 3: Robustness frontiers. **A**: Dual-end utility tracks local mutation-rank fidelity, while the priority-shuffle control preserves sequence AUROC but destroys local agreement. **B**: Sampling-only control improves with denoising steps and guide emphasis but saturates below the fixed full-method reference. **C**: RF-only optimization attains high proxy scores alongside poor cross-evaluator utility and high frozen-DPLM perplexity. Marker area encodes valid-and-joint yield.

6 Conclusion

We introduced **PPDiff**, which uses an external property prior to coordinate fixed-budget training corruption and readiness-aware sampling while retaining the MDM architecture. Across multiple peptide properties and targets, as well as diverse evaluation frameworks, the shared prior improves control most when used at both stages, while factorial and shuffled-prior controls support coordination of reconstruction practice and property-directed refinement.

The target-specific candidates and docking poses are computational hypotheses requiring experimental validation. Next steps include extending the representation to non-canonical amino acids and validating prioritized candidates through wet-lab assays, including binding, toxicity, stability, and functional evaluations.

AI use statement

For this manuscript revision, we used OpenAI Codex solely to check grammar and assist with LaTeX formatting. It was not used to generate research data, design or run experiments, implement the reported method, formulate mathematical claims, assist with proofs, or compute the reported results.

All AI-assisted edits were checked against the source manuscript and reported tables. The authors take full responsibility for the final text, claims, and artifacts.

Ethics statement

This study uses public peptide and protein resources and computational models. It involves no human participants, animals, clinical records, or private personal data. Any synthesis or deployment requires target-specific experimental validation, toxicity and off-target assessment, and appropriate institutional biosafety review. Appendix A documents dataset provenance, Appendix G describes the computational screen, and Appendix H discusses dual-use implications.

Reproducibility statement

Section 3 and Appendix B.1 specify the method and reference sampler. Appendix A documents data construction, model roles, baselines, metrics, seed hierarchies, and statistical tests. The remaining appendices report mechanism controls, quality and evaluator audits, computational costs, and target-specific screening procedures. The anonymized code repository and configuration files are at `https://anonymous.4open.science/r/PPDiff-3F4C`.

References

Piyush Agrawal, Sherry Bhalla, Salman Sadullah Usmani, Sandeep Singh, Kumardeep Chaudhary, Gajendra P. S. Raghava, and Ankur Gautam. CPPsite 2.0: A repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Research*, 44(D1):D1098–D1103, 2016. doi: 10.1093/nar/gkv1266.

Sarah Alamdari, Nitya Thakkar, Rianne van den Berg, Alex X. Lu, Nicolo Fusi, Ava P. Amini, and Kevin K. Yang. Protein generation with evolutionary diffusion: Sequence is all you need. *bioRxiv*, 2023. doi: 10.1101/2023.09.11.556673. URL `https://github.com/microsoft/evodiff`.

Jacob Austin, Daniel D. Johnson, Jonathan Ho, Daniel Tarlow, and Rianne van den Berg. Structured denoising diffusion models in discrete state-spaces. In *Advances in Neural Information Processing Systems*, volume 34, pp. 17981–17993, 2021. URL `https://proceedings.neurips.cc/paper/2021/hash/958c530554f78bcd8e97125b70e6973d-Abstract.html`.

Rainier Barrett, Shaoyi Jiang, and Andrew D. White. Classifying antimicrobial and multifunctional peptides with bayesian network models. *Peptide Science*, 110(4):e24079, 2018. doi: 10.1002/pep2.24079.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

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

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

Gabriel Cabas-Mora, Anamaría Daza, Nicole Soto-García, Valentina Garrido, Diego Alvarez, Marcelo Navarrete, Lindybeth Sarmiento-Varón, Julieta H. Sepúlveda Yañez, Mehdi D. Davari, Frederic Cadet, Álvaro Olivera-Nappa, Roberto Uribe-Paredes, and David Medina-Ortiz. Peptipedia v2.0: A peptide sequence database and user-friendly web platform. A major update. *Database*, 2024: baae113, 2024. doi: 10.1093/database/baae113.

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*, 44 (6):1002–1010, 2026. doi: 10.1038/s41587-025-02761-2. Published online 13 August 2025.

Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. AutoDock Vina 1.2.0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling*, 61(8):3891–3898, 2021. doi: 10.1021/acs.jcim.1c00203.

Ahmed Elnaggar, Michael Heinzinger, Christian Dallago, Ghalia Rehawi, Yu Wang, Llion Jones, Tom Gibbs, Tamas Feher, Christoph Angerer, Martin Steinegger, et al. ProtTrans: Toward understanding the language of life through self-supervised learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7112–7127, 2022. doi: 10.1109/TPAMI.2021.3095381.

Forli Lab. Meeko: Interface for autodock, 2026. URL https://github.com/forlilab/Meeko. Software, version 0.7.1. Accessed September 16, 2026.

Ankur Gautam, Harinder Singh, Atul Tyagi, Kumardeep Chaudhary, Rahul Kumar, Pallavi Kapoor, and Gajendra P. S. Raghava. CPPsite: A curated database of cell penetrating peptides. *Database*, 2012:bas015, 2012. doi: 10.1093/database/bas015.

Ankur Gautam, Kumardeep Chaudhary, Sandeep Singh, Anshika Joshi, Priya Anand, Abhishek Tuknait, Deepika Mathur, Grish C. Varshney, and Gajendra P. S. Raghava. Hemolytik: A database of experimentally determined hemolytic and non-hemolytic peptides. *Nucleic Acids Research*, 42 (D1):D444–D449, 2014. doi: 10.1093/nar/gkt1008.

Shrey Goel, Peregrine M. Schray, Yinuo Zhang, Sophia Vincoff, Huong T. Kratochvil, and Pranam Chatterjee. Token-level guided discrete diffusion for membrane protein design. *arXiv preprint arXiv:2410.16735*, 2024. doi: 10.48550/arXiv.2410.16735. Revised September 2025.

Nate Gruver, Samuel Stanton, Nathan C. Frey, Tim G. J. Rudner, Isidro Hotzel, Julien Lafrance-Vanasse, Arvind Rajpal, Kyunghyun Cho, and Andrew Gordon Wilson. Protein design with guided discrete diffusion. In *Advances in Neural Information Processing Systems*, volume 36, pp. 12489–12517, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/29591f355702c3f4436991335784b503-Abstract-Conference.html.

Chakradhar Guntuboina, Adrita Das, Parisa Mollaei, Seongwon Kim, and Amir Barati Farimani. PeptideBERT: A language model based on transformers for peptide property prediction. *The Journal of Physical Chemistry Letters*, 14(46):10427–10434, 2023. doi: 10.1021/acs.jpclett.3c02398.

Emiel Hoogeboom, Didrik Nielsen, Priyank Jaini, Patrick Forré, and Max Welling. Argmax flows and multinomial diffusion: Learning categorical distributions. In *Advances in Neural Information Processing Systems*, volume 34, pp. 12454–12465, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/67d96d458abdef21792e6d8e590244e7-Abstract.html.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Kimia Kardani and Azam Bolhassani. Cppsite 2.0: An available database of experimentally validated cell-penetrating peptides predicting their secondary and tertiary structures. *Journal of Molecular Biology*, 433(11):166703, 2021. doi: 10.1016/j.jmb.2020.11.002.

Wouter Kool, Herke van Hoof, and Max Welling. Stochastic beams and where to find them: The Gumbel-top-k trick for sampling sequences without replacement. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3499–3508, 2019. URL https://proceedings.mlr.press/v97/kool19a.html.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023. doi: 10.1126/science.ade2574.

Linrui Ma, Yufei Cui, Kai Han, and Yunhe Wang. Mask is what DLLM needs: A masked data training paradigm for diffusion LLMs. *arXiv preprint arXiv:2603.15803*, 2026. doi: 10.48550/arXiv.2603.15803.

Dorian Migoń, Maciej Jaśkiewicz, Damian Neubauer, Marta Bauer, Emilia Sikorska, Elżbieta Kamysz, and Wojciech Kamysz. Alanine scanning studies of the antimicrobial peptide aurein 1.2. *Probiotics and Antimicrobial Proteins*, 11(3):1042–1054, 2019. doi: 10.1007/s12602-018-9501-0.

Markus Muttenthaler, Glenn F. King, David J. Adams, and Paul F. Alewood. Trends in peptide drug discovery. *Nature Reviews Drug Discovery*, 20(4):309–325, 2021. doi: 10.1038/s41573-020-00135-8.

Surag Nair, Avanti Shrikumar, Jacob Schreiber, and Anshul Kundaje. fastISM: Performant in silico saturation mutagenesis for convolutional neural networks. *Bioinformatics*, 38(9):2397–2403, 2022. doi: 10.1093/bioinformatics/btac135.

Shen Nie, Fengqi Zhu, Zebin You, Xiaolu Zhang, Jingyang Ou, Jun Hu, Jun Zhou, Yankai Lin, Ji-Rong Wen, and Chongxuan Li. Large language diffusion models. *arXiv preprint arXiv:2502.09992*, 2025. doi: 10.48550/arXiv.2502.09992.

Hunter Nisonoff, Junhao Xiong, Stephan Allenspach, and Jennifer Listgarten. Unlocking guidance for discrete state-space diffusion and flow models. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=XsgHl54yO7.

OpenMM contributors. PDBFixer: Fixing problems in protein data bank files, 2026. URL https://github.com/openmm/pdbfixer. Software, version 1.12. Accessed September 16, 2026.

Fred Zhangzhi Peng, Zachary Bezemek, Sawan Patel, Jarrid Rector-Brooks, Sherwood Yao, Alexander Tong, and Pranam Chatterjee. Path planning for masked diffusion models with applications to biological sequence generation. In *Generative and Experimental Perspectives for Biomolecular Design Workshop at ICLR*, 2025. URL https://openreview.net/forum?id=VOW9YCgmMA.

Fred Zhangzhi Peng, Zachary Bezemek, Jarrid Rector-Brooks, Shuibai Zhang, Anru R. Zhang, Michael Bronstein, Alexander Tong, and Avishek Joey Bose. Planner aware path learning in diffusion language models training. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=lAlI5FuIf7. Oral presentation. arXiv:2509.23405.

Jarrid Rector-Brooks, Mohsin Hasan, Zhangzhi Peng, Zachary Quinn, Chenghao Liu, Sarthak Mittal, Nouha Dziri, Michael Bronstein, Yoshua Bengio, Pranam Chatterjee, Alexander Tong, and Avishek Joey Bose. Steering masked discrete diffusion models via discrete denoising posterior prediction. In *International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=Ombm8S40zN.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5):742–754, 2010. doi: 10.1021/ci100050t.

Subham Sekhar Sahoo, Marianne Arriola, Yair Schiff, Aaron Gokaslan, Edgar Marroquin, Justin T. Chiu, Alexander Rush, and Volodymyr Kuleshov. Simple and effective masked diffusion language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 130136–130184, 2024. doi: 10.52202/079017-4135.

Jiaxin Shi, Kehang Han, Zhe Wang, Arnaud Doucet, and Michalis K. Titsias. Simplified and generalized masked diffusion for discrete data. In *Advances in Neural Information Processing Systems*, volume 37, pp. 103131–103167, 2024. doi: 10.52202/079017-3277.

Ayushi Singh, Kavin Raj SA, Anand Singh Rathore, and Gajendra P. S. Raghava. Hemolytik 2: An updated database of hemolytic peptides and proteins. *Chemical Research in Toxicology*, 39(2):208–215, 2026. doi: 10.1021/acs.chemrestox.5c00322.

Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017. doi: 10.1038/nbt.3988.

Paulina Szymczak, Marcin Możejko, Tomasz Grzegorzek, Radosław Jurczak, Marta Bauer, Damian Neubauer, Karol Sikora, Michał Michalski, Jacek Sroka, Piotr Setny, Wojciech Kamysz, and Ewa Szczurek. Discovering highly potent antimicrobial peptides with deep generative model HydrAMP. *Nature Communications*, 14(1):1453, 2023. doi: 10.1038/s41467-023-36994-z.

Sophia Tang, Yinuo Zhang, and Pranam Chatterjee. PepTune: De novo generation of therapeutic peptides with multi-objective-guided discrete diffusion. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 59017–59065, 2025. URL https://proceedings.mlr.press/v267/tang25n.html.

The UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617, 2025. doi: 10.1093/nar/gkae1010.

Oleg Trott and Arthur J. Olson. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: 10.1002/jcc.21334.

Shuzhe Wang, Jagna Witek, Gregory A. Landrum, and Sereina Riniker. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *Journal of Chemical Information and Modeling*, 60(4):2044–2058, 2020. doi: 10.1021/acs.jcim.0c00025.

Xinyou Wang, Zaixiang Zheng, Fei Ye, Dongyu Xue, Shujian Huang, and Quanquan Gu. Diffusion language models are versatile protein learners. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 52309–52333, 2024. URL https://proceedings.mlr.press/v235/wang24ct.html.

Andrew D. White, Ann K. Nowinski, Wenjun Huang, Andrew J. Keefe, Fang Sun, and Shaoyi Jiang. Decoding nonspecific interactions from nature. *Chemical Science*, 3(12):3488–3494, 2012. doi: 10.1039/C2SC21135A.

Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. doi: 10.2307/3001968.

Jiahui Zhang, Rouyi Wang, Kuangqi Zhou, Tianshu Xiao, Lingyan Zhu, Yaosen Min, and Yang Wang. PepBenchmark: A standardized benchmark for peptide machine learning. In *International Conference on Learning Representations*, 2026. URL https://arxiv.org/abs/2604.10531. arXiv:2604.10531.

A Experimental Protocol and Reproducibility

This appendix consolidates the data construction, separation between guides and evaluators, comparison scope, statistics, and reproducibility details summarized in Section 4.1.

A.1 Properties, data construction, and generator

We study permeability, hemolysis safety, and nonfouling, with one RF-based ISM guide per property. Permeability positives come from experimentally validated cell-penetrating-peptide records in CPPsite and CPPsite 2.0 (Gautam et al., 2012; Agrawal et al., 2016; Kardani & Bolhassani, 2021). Hemolysis labels come from experimentally determined hemolytic and non-hemolytic records in Hemolytik (Gautam et al., 2014; Singh et al., 2026), supplemented by nearly 200 records from Peptipedia v2.0 (Cabas-Mora et al., 2024). The nonfouling set follows Barrett et al. (2018), based on the analysis of 1,162 protein surfaces and chaperone interiors by White et al. (2012). Selected peptide mimics were experimentally tested, while the larger sequence set was derived from surface fragments and composition principles. Its labels are therefore partly rule-constructed.

After MMseqs2 de-duplication (Steinegger & Söding, 2017), the datasets contain approximately 8,500 permeability, 13,500 hemolysis, and 5,800 nonfouling peptides. We remove redundancy above 90% identity before splitting and combine 30% identity clusters with 3-mer leakage control, following

Table 9: Data-construction parameters.

Item	Value
MMseqs2 identity / coverage	0.90 / 0.80
Before / after de-duplication	perm. $\sim$12k$\rightarrow$8.5k, hemo. $\sim$17k$\rightarrow$13.5k, and nonfoul. $\sim$7.2k$\rightarrow$5.8k
Weak-negative pool	UniProt unlabeled + reversed, excluding AMP/CPP/toxin
Positive/negative matching	JS ≤ 0.05 for length, charge, hydrophobicity, AA, dipeptide
Split	hybrid: 30% identity + 3-mer leakage control
Enriched motif control	FDR < 0.05, odds ratio < 2.0

PepBenchmark's preprocessing principles (Zhang et al., 2026). Table 9 summarizes the construction thresholds.

Identity filtering reduces near-duplicate leakage, while distribution matching limits shortcuts based on length, charge, hydrophobicity, or amino-acid composition.

The backbone is DPLM-150M initialized from a common short-peptide LoRA checkpoint (Wang et al., 2024; Hu et al., 2022). Every property-aware run starts from this checkpoint and preserves its architecture. Generation lengths are sampled from the task distribution and held fixed across methods. Unless noted, each condition contains 1,000 raw generations and uses 64 denoising iterations, Gumbel-argmax token sampling, and seed 42 for point estimates. RF-based sampling guidance uses scorer warm-up 0.1 and refresh interval $M = 5$; gradient guidance is evaluated every step (Appendix A.3). Appendix A.5 specifies the seed hierarchies.

A.2 Guide and evaluators

RF-based ISM guides the three sequence properties, while frozen PepMLM supplies the target-conditioned guide (Section 3.3). Separate models evaluate the generated sequences: PepBenchmark's ESM-150M-F, based on ESM2-150M (Lin et al., 2023; Zhang et al., 2026), provides headline scores; PepBenchmark's ECFP fingerprint model (Rogers & Hahn, 2010) and ProtBert-based PeptideBERT (Elnaggar et al., 2022; Guntuboina et al., 2023) provide cross-checks. We train a PeptideBERT permeability head because the original model lacks one. AutoDock Vina supplies the computational binding screen (Trott & Olson, 2010; Eberhardt et al., 2021). Table 10 summarizes these roles.

Table 10: Model roles: RF and PepMLM guide generation; ESM, fingerprint, and PeptideBERT evaluate properties; Vina screens frozen candidate pools.

Role	Model	Input repr.	Separation from guide	Use
Property guide	sequence RF + ISM	padded one-hot AA	steering signal only	train mask + sample commit
Binding guide	PepMLM-650M	target + peptide AA	steering signal only	target-specific prior
Primary eval.	ESM-150M-F	pretrained PLM	architecture, provenance, cluster split	headline property scores
Cross-check	PepBenchmark FP	ECFP6, 1,024 bit	feature space and provenance	common-evaluator means
Cross-check	PeptideBERT	ProtBert	architecture with a custom permeability head	threshold pass rates
Docking screen	AutoDock Vina	prepared 3D structures	no generation-time query	frozen-pool screen

The guide training set has no exact duplicates or 90% identity cluster overlap with the primary-evaluator test set. Appendix E.1 quantifies overlap at looser thresholds. Macro-averaged held-out AUCs range from 0.783 to 0.835, and expected calibration error is at most 0.0634 (Appendix E.3).

We assess RF guide reliability on the homology- and 3-mer-controlled held-out split using AUROC, PR-AUC, MCC, and expected calibration error. To evaluate the local rankings used by **PPDIFF**, we enumerate each held-out peptide's $19L$ single-residue alternatives and compare RF mutation effects with the separate evaluator ensemble using mean within-sequence Spearman correlation and top-20% position overlap. We then rerun the dual-end pipeline with 100%, 50%, 25%, and 10% of guide training data, 40% label noise, fully shuffled labels, and within-sequence shuffled priorities, holding all other optimization and generation budgets fixed (Appendix F.1).

A.3 BASELINES AND COMPARISON SCOPE

We compare with the property-neutral backbone checkpoint, continued uniform-mask training, a within-sequence shuffled-prior control, PAPL with its planner-aware objective (Peng et al., 2026), gradient classifier guidance, and PepTune (Tang et al., 2025). The matched PAPL comparison retains the PAPL-trained checkpoint and replaces only its inference operator with our `improvable_remask` sampler and RF-based ISM P/S/N priorities.

Gradient classifier guidance. This inference-only baseline uses the property-neutral DPLM checkpoint and one differentiable classifier $C_{\psi,p}$ per property, trained on the corresponding RF guide data and splits. Each classifier takes 20-channel amino-acid inputs and uses two 1D convolutional layers with 128 channels each, GELU activations, masked mean pooling over residue positions, and a linear binary head with sigmoid output. The heads predict the desired property, including non-hemolysis for safety, and are separate from all reporting evaluators. Classifier and backbone parameters remain fixed during generation. Guidance changes amino-acid proposal logits while retaining the base DPLM remaining-mask schedule and state-update rule.

Let $\ell_{t,i} \in \mathbb{R}^{20}$ be the canonical-residue logits and $\mathcal{M}_{t-1}$ the currently masked residue positions. The classifier input combines soft distributions at masked sites with retained one-hot residues:

$$Z_{t,i} = \begin{cases} \mathrm{softmax}(\ell_{t,i}/\tau_c), & i \in \mathcal{M}_{t-1}, \\ \mathrm{onehot}(x^{\mathrm{state}}_{t-1,i}), & i \notin \mathcal{M}_{t-1}, \end{cases} \qquad R_\psi(Z_t) = \sum_{p \in \mathcal{P}} w_p \log\big(C_{\psi,p}(Z_t) + \epsilon\big). \tag{12}$$

Here $\tau_c > 0$ is the relaxation temperature, $\epsilon > 0$ is a numerical stabilizer, and $w_P = w_S = w_N = 1/3$. We detach the DPLM outputs and treat the logits as leaf variables, backpropagating through the soft relaxation and frozen classifiers but not through DPLM parameters. For each masked position, the chain rule gives

$$G_{t,i} = \nabla_{\ell_{t,i}} R_\psi(Z_t) = \frac{1}{\tau_c}\left[\mathrm{diag}(Z_{t,i}) - Z_{t,i} Z_{t,i}^\top\right] \nabla_{Z_{t,i}} R_\psi(Z_t), \tag{13}$$

$$\bar{G}_{t,i} = \frac{G_{t,i}}{\|G_{t,i}\|_2 + \epsilon}, \qquad \widetilde{\ell}_{t,i} = \ell_{t,i} + \lambda_c \bar{G}_{t,i}, \quad \lambda_c \geq 0. \tag{14}$$

The resulting logit gradient is L2-normalized over the 20 amino-acid components at each position after combining property gradients. Thus λ_c controls the logit perturbation, whereas τ_c controls the soft relaxation. Gumbel-argmax sampling uses $\widetilde{\ell}_{t,i}$ to generate proposals, followed by the base state update.

Guidance starts at the first step. Each of the $T = 64$ steps evaluates the three property heads and computes the combined reward's logit gradient, giving 64 denoiser evaluations and 64 guidance rounds (Table 23).

No-denoiser and reranking controls. To test whether the RF can replace the denoiser, we add five controls. Uniform-AA and empirical-unigram random sampling use the fixed requested length multiset but no sequence model. Random+RF top-K ranks random candidates under the equal-weight three-guide objective. RF-only greedy coordinate ascent starts from random sequences, enumerates all 19 substitutions at each position, accepts the largest objective-improving mutation, and uses random restarts. Both RF-only optimizers are matched to the RF-call budget of guided **PPDIFF** sampling. DPLM+RF reranking instead generates 64 complete backbone candidates and selects them only after generation. These controls distinguish unconstrained proxy optimization, proposal quality from the peptide model, and trajectory-level use of the prior. Appendix F.3 specifies their measured transfer, quality, and query costs.

PepTune. PepTune uses a trained SMILES-MDLM generator with MCTG guidance. We retain canonical outputs, resampling until the target count is reached and selecting solely by canonicality. Valid denotes canonical-protocol yield under our residue-level evaluation. The fixed canonical-output set is scored by the same ESM-150M-F, fingerprint, and PeptideBERT evaluators used for other methods. PepTune's native generated-set PPL is 2.598 under its SPE/SMILES model. Its vocabulary, token units, and evaluation population differ from residue-model validation PPL, so these likelihood statistics are reported separately.

A.4 METRICS AND STATISTICAL TESTS

Property control is measured by evaluator means and distribution shift from the start model. A sequence is *scoreable* if it contains only canonical residues, has no unresolved mask or special token, and lies inside the prespecified length support. Property statistics use this population, while the stricter protocol Valid metric is reported separately. Appendix C reports Jensen-Shannon (JS) divergence, 1-Wasserstein distance, and two-sample Kolmogorov-Smirnov (KS) tests. JS uses 50 fixed bins on $[0, 1]$, base-2 logarithms, and 10^{-12} additive smoothing. Wasserstein and KS use unbinned scores. Inferential interpretation applies the Benjamini and Hochberg procedure across the 21 combinations of policies and properties. For evaluator o, the balanced marginal utility is

$$J_o(x; \mathbf{w}) = \sum_{p \in \mathcal{P}} w_p \, s_{o,p}(x), \quad w_p \geq 0, \quad \sum_p w_p = 1, \tag{15}$$

where every $s_{o,p} \in [0, 1]$ is oriented so that larger is better. The headline setting uses equal weights. Joint satisfaction and worst-property performance are measured by

$$H_o(x) = \mathbb{I}[s_{o,P}(x) \geq \tau_{o,P},\ s_{o,S}(x) \geq \tau_{o,S},\ s_{o,N}(x) \geq \tau_{o,N}], \tag{16}$$

$$m_o(x) = \min\{s_{o,P}(x), s_{o,S}(x), s_{o,N}(x)\}. \tag{17}$$

Each $\tau_{o,p}$ is fixed before generation by maximizing Matthews correlation coefficient on the corresponding validation split. We report 95% sequence-bootstrap intervals for single-seed headline means; these quantify sequence-level uncertainty at a fixed checkpoint. The two-factor experiment crosses training priority with guided sampling, and its interaction is

$$\mathcal{I} = J_{\text{both}} - J_{\text{train}} - J_{\text{sample}} + J_{\text{base}}. \tag{18}$$

Interaction intervals are obtained by bootstrapping seed means within each cell.

For the stress tests, let V denote the protocol-valid indicator. We report $\Pr(V = 1, H_o = 1)$ over all raw generations and the yield-adjusted utility $\Pr(V = 1)\,\mathbb{E}[J_o \mid \text{scoreable}]$. Generated-sequence PPL_{gen} is the frozen-DPLM NELBO perplexity on outputs; checkpoint validation PPL uses the held-out validation set (Table 28). Composition OOD uses the training-set 99th percentile of amino-acid-composition Mahalanobis distance. We also audit longest homopolymer runs, maximum single-residue fraction, denoiser NFE, and RF forward-equivalent calls. Inference accounting excludes fine-tuning and prior-precomputation costs.

Generation quality uses protocol Valid, uniqueness, pairwise diversity, nearest-training-neighbor identity, and masked-diffusion validation perplexity (Table 28). A protocol-valid output must be scoreable, contain no run longer than five residues, and contain no residue occupying more than half of the sequence. Uniqueness uses protocol-valid outputs; diversity and nearest-neighbor identity use scoreable outputs. The original docking screen uses five RDKit ETKDGv3/MMFF conformers (Wang et al., 2020) and a top-20 pool frozen by the PepMLM-based reward and P/S/N feasibility before any Vina call. Each candidate is paired with an unrestricted composition-preserving scramble and a charge-pattern shuffle. Their mean ligand efficiency (LE) defines its control. We apply a paired Wilcoxon signed-rank test to the 20 candidate-minus-control differences (Wilcoxon, 1945), followed by Benjamini–Hochberg false-discovery-rate correction across seven targets (Benjamini & Hochberg, 1995).

A.5 REPRODUCIBILITY NOTES

The primary interaction analysis uses five generation seeds per cell, one fixed checkpoint per training level, and 1,000 raw sequences per replicate. The checkpoint-robustness analysis uses three independent base/fine-tuning seeds with three generation seeds per checkpoint. The matched intact-versus-shuffled factorial and guide-degradation studies each use five independent fine-tuning seeds with three generation seeds per checkpoint. The sampling-only frontier uses matched generation seeds across its $T \times \lambda_g$ grid.

All default DPLM-family sampling comparisons use 64 denoiser evaluations and the same requested length multiset. Guided **PPDIFF** and PAPL+matched use 13 RF-based ISM refreshes. The sampling-only grid crosses $T \in \{32, 64, 128, 256\}$ with $\lambda_g \in \{0.5, 1, 2, 4\}$, using $s_i(t) = g_i(t)^{\lambda_g} r_i(t)$. Hyperparameters are selected on the prespecified validation protocol, while the full method remains

fixed at $T = 64, \lambda_g = 1$. For each target-specific fine-tuning run, PepMLM scores cover 50% of training examples to control scorer cost; P/S/N priorities cover the complete training set. Target scoring is applied throughout sampling. The docking reconstructions document receptor preparation, boxes, software versions, conformer seeds, and run manifests.

B Method Specification and Mechanism Controls

This section specifies the reference implementation, positions its training and sampling rules relative to existing approaches, and reports mechanism controls and computational costs.

Table 11: Mechanistic positioning of trajectory-control methods. "External" denotes a task-property model separate from the denoiser.

Method family	Train control	Sample control	External task-property guide
Classifier/token guidance	None	gradient/token	classifier
PepTune (MCTG)	None	tree search	reward ensemble
PAPL	planned path	planned path	none
Information-density masking	priority mask	None	heuristic/LLM
PPDiff	priority mask	importance × readiness	RF-based ISM, target-conditioned PepMLM

Table 12: Default **PPDiff** configuration, preserving the denoiser and token-head architecture.

Component	Choice	Default
Backbone	DPLM-150M + peptide LoRA	same architecture
Prior	RF-based ISM and PepMLM for binding	weighted rank fusion
Priority	$q_i = c_i - \lambda_u u_i$	$\lambda_u = 0.5$
Headline train policy	deterministic top-M_t priority	ρ, w, f: not used
Stochastic ablations	Gumbel top-M_t	$\rho_{\max} = 0.7$, $w = 2$, $f = 0.4$
Commit score	importance × readiness	$(\lambda_g, \alpha, \beta, \gamma) = (1, 1, 1, 1)$
Sampler	improvable remask	$T = 64$, $M = 5$

The headline policy uses the continuous priority ranking directly. Parameters ρ, w, and f apply to stochastic or binary ablations; Appendix B.1 specifies their operators and the trainable parameters.

B.1 Reference implementation

The uncertainty estimator, rewrite rule, and update order below specify the reported sampler.

B.1.1 Uncertainty, ranks, and reconstruction exposure

For tree b of property guide p, compute its own wild-type-referenced criticality,

$$c_{i,b}^{(p)} = -\frac{1}{19} \sum_{a \neq x_i} \left[f_{p,b}(x^{i \to a}) - f_{p,b}(x) \right], \tag{19}$$

$$\nu_i^{(p)} = \mathrm{Std}_b \left(c_{i,b}^{(p)} \right), \qquad u_i^{(p)} = \frac{\nu_i^{(p)} - \min_j \nu_j^{(p)}}{\max_j \nu_j^{(p)} - \min_j \nu_j^{(p)}}. \tag{20}$$

The population standard deviation uses at most 32 trees, selected by a deterministic stride through the forest. If the denominator is at most 10^{-12}, all uncertainties are zero. Setting $\lambda_u = 0$ skips these additional tree passes. A deterministic non-ensemble guide, including the target-conditioned scorer, has $u_i = 0$.

Percentile ranks use averaged ties and are divided by the number of residue positions: $\mathrm{PctRank}(q)_i = \mathrm{rank}_{\mathrm{avg}}(q_i)/L$, with rank one the lowest. The $L = 1$ special case returns $1/2$. A single property uses its raw uncertainty-discounted criticality; multiple properties use Equation 4. Retaining multiple guide objects therefore applies rank fusion even at a single-objective weight endpoint.

Let ϱ_i be the descending priority rank. Under deterministic corruption, the reconstruction-exposure probability is

$$E_i^{\text{train}} = \Pr(M_t \geq \varrho_i). \tag{21}$$

For mask counts uniform on $\{1, \ldots, L\}$, this is $(L - \varrho_i + 1)/L$. The reference configuration samples discrete diffusion time $t \in \{1, \ldots, D\}$ with $D = 500$ and linear loss weight $a(t) = (D - t + 1)/D$; the constant-weight option has $a(t) = 1$. Weighted exposure is $\mathbb{E}_t[a(t)\mathbb{I}(M_t \geq \varrho_i)]$. Token loss is normalized by the batch's non-padding token count.

The ordered policy uses continuous priorities directly. The continuous `weighted_priority` ablation samples without replacement with weights proportional to $(\text{MinMax}(q_i) + \epsilon)^\beta$ ($\beta = 2$, $\epsilon = 10^{-6}$); the binary/curriculum family uses Equation 7. The reference fine-tuning configuration updates rank-16 LoRA adapters with dropout 0.1, embeddings, and the language-model head, preserving the architecture.

B.1.2 Signed improvement signal and the rewrite quota

The headline operator retains a signed signal separately from the percentile-fused priority:

$$h_i(\hat{x}_t) = \sum_p w_p \bar{\Delta}_i^{(p)}(\hat{x}_t). \tag{22}$$

This signed mean combines raw, oriented guide-score changes, whereas commitment priorities use percentile ranks. Positive h_i identifies positions with beneficial substitutions on average for another denoising attempt; the denoiser supplies the next proposal. The raw-score combination remains sensitive to relative guide scales.

Use increasing generation-step indices $t = 1, \ldots, T$ and let $\mathcal{U}$ be the mutable residue positions, excluding special and externally fixed tokens. The default linear schedule gives nominal remaining-mask budget

$$k_t = \lfloor |\mathcal{U}|(1 - t/T) \rfloor. \tag{23}$$

The eligibility set and effective forced quota are

$$\mathcal{E}_t = \{i \in \mathcal{U} : h_i > 0\}, \tag{24}$$

$$Q_t = \begin{cases} \min\left\{|\mathcal{E}_t|, \max(1, \lfloor \eta k_t \rfloor)\right\}, & k_t > 0, \\ 0, & k_t = 0, \end{cases} \qquad \eta = 0.7, \tag{25}$$

$$\mathcal{F}_t = \text{TopK}_{i \in \mathcal{E}_t}\{h_i; Q_t\}. \tag{26}$$

Eligibility includes both masked and unmasked sites. Selection can therefore keep a site masked or return an already revealed site to the mask state. There is no per-position lifetime cap or additional global rewrite counter in this operator. The minimum-one convention means that setting `max_forced_ratio` to zero does not disable forcing when candidates exist and $k_t > 0$. The no-forcing ablation uses the `joint_readiness` strategy without the override.

For nonempty $\mathcal{F}_t$, its commitment scores are replaced by

$$\widetilde{s}_i = \begin{cases} \min_{j \in \mathcal{U} \setminus \mathcal{F}_t} s_j - 1, & i \in \mathcal{F}_t, \\ s_i, & \text{otherwise.} \end{cases} \tag{27}$$

An empty survivor set uses zero as the floor before subtracting one. The sampler forms $\mathcal{B}_t$ from scores strictly below the $(k_t + 1)$-st smallest score. Distinct scores select the lowest k_t positions; cutoff ties can yield fewer. Forced selection uses the tensor library's top-k tie behavior. Gumbel noise is used for residue proposals only.

B.1.3 Complete update order

Let $\mathcal{M}_{t-1}$ be the current masked set and x_{t-1}^{state} the retained partial sequence. The backbone predicts distributions at all mutable positions and samples a full proposal $\hat{x}_t$. The guide scores that proposal, which can differ from the tokens currently retained at unmasked positions. The state update is

$$x_{t,i}^{\text{state}} = \begin{cases} \mathbf{m}, & i \in \mathcal{B}_t, \\ \hat{x}_{t,i}, & i \in \mathcal{M}_{t-1} \setminus \mathcal{B}_t, \\ x_{t-1,i}^{\text{state}}, & \text{otherwise,} \end{cases} \qquad \mathcal{M}_t = \mathcal{B}_t. \tag{28}$$

Previously unmasked positions outside $\mathcal{B}_t$ keep their retained tokens. Newly remasked positions use the next iteration's denoiser call. At $t = T$, $k_t = 0$ and the mask set is empty.

Table 13: Reference sampling algorithm. Step indices increase from the all-masked state to the final sequence.

	Prior-guided sampling with signed-mean remasking
1	Initialize all mutable residues to mask, the cached prior to empty, and previous probabilities to empty.
2	For $t = 1, \ldots, T$, make one backbone call and sample the full proposal $\hat{x}_t$.
3	Compute confidence, margin, and stability from the categorical distributions. Use stability one initially; otherwise use $1 - \mathrm{JSD}_2$ against the preceding iteration.
4	If $(t-1) \bmod M = 0$, refresh (q, h) on $\hat{x}_t$; otherwise reuse the cached values.
5	For the first $\lfloor 0.1T \rfloor$ steps use the raw decoder ranking. Subsequently compute $s_i = g_i^{\lambda_g} r_i$, then apply Equations 22–26.
6	Form $\mathcal{B}_t$ from the modified ranking and update the state by Equation 28. Save the current probabilities for the next step.
7	Return the final retained sequence.

Guide evaluations occur during warm-up even though guidance is not yet applied. Thus $T = 64$, $M = 5$ refreshes at steps $1, 6, \ldots, 61$, giving 13 refreshes. Previous probability vectors are retained after remasking. Readiness factors are floored at 10^{-6} before exponentiation, and JSD computation uses a 10^{-9} numerical floor.

B.1.4 The conservative operator and additive normalization

The `weighted_T_V` alternative uses raw decoder scores ℓ_i and mutation-effect variance $v_i = \sum_p w_p \mathrm{Var}_a(\Delta_{i,a}^{(p)})$. With within-sequence min–max normalization N over mutable positions, it ranks by

$$s_i^{\mathrm{WTV}} = a_t N(\ell_i) + (1 - a_t) N(q_i) - N(v_i), \qquad a_t = 0.9 - 0.6t/T. \tag{29}$$

Here v_i measures variation across amino-acid substitutions; u_i measures disagreement across RF trees. The normalizer uses denominator $\max - \min + 10^{-8}$ and maps a singleton to $1/2$.

In the additive ablation, $z(a_i) = (a_i - \mu_{\mathcal{U}}(a))/(\mathrm{Std}_{\mathcal{U}}(a) + 10^{-8})$, recomputed per sequence and step over mutable positions with the sample standard deviation. Singleton sets map to zero.

B.2 Independent-checkpoint and prior-integrity controls

Table 15 gives the complete point-estimate contrasts. Table 6 evaluates their checkpoint variability, complementing the generation-seed analysis in Table 5.

B.3 Direction of the property preference

Let T^+ denote training that masks higher-priority positions and T^- training that masks lower-priority positions. Similarly, S^+ and S^- denote sampling that favors the higher and lower ends of the property ranking. Table 16 compares these four combinations.

The utility differences $T^+S^+ - T^+S^-$ and $T^+S^+ - T^-S^+$ are 0.2047 and 0.1462. Their difference of differences is approximately 0.0697 using the rounded utility column, consistent with emphasizing property-relevant residues at both stages. Valid ranges from 0.849 to 0.858, while the strongest control has the highest generated-sequence PPL, indicating a control–naturalness tradeoff.

B.4 Remasking operator and rewrite budget

Table 17 separates commitment scoring from additional rewrite opportunities. The two no-remask controls retain revealed residues. For the improvement-ranked comparison, the best-mutant gain is

$$d_i = \max_{a \neq \hat{x}_i} \left[\sum_p w_p \Delta_{i,a}^{(p)} \right]_+ . \tag{30}$$

Table 14: Independent-checkpoint robustness. Property columns show mean ± SD over three independent base/fine-tuning seeds, each decoded with three generation seeds. Balanced utility reports means.

Method	Permeability	Safety	Nonfouling	Balanced utility
Backbone start	0.5998 ± 0.0102	0.6237 ± 0.0113	0.6585 ± 0.0121	0.6273
PPDIFF training only	0.7003 ± 0.0188	0.6984 ± 0.0172	0.7192 ± 0.0207	0.7060
PPDIFF sampling only	0.7516 ± 0.0112	0.7674 ± 0.0104	0.7730 ± 0.0123	0.7640
PAPL + matched sampler	0.8165 ± 0.0218	0.8074 ± 0.0193	0.8025 ± 0.0237	0.8088
PPDIFF full	$\mathbf{0.8998 \pm 0.0158}$	$\mathbf{0.8975 \pm 0.0176}$	$\mathbf{0.9084 \pm 0.0152}$	**0.9019**

Table 15: 2×2 point-estimate controls on balanced utility. Table 5 reports the intact-prior interaction with generation-seed uncertainty.

Position prior	Start	Train only	Sample only	Both	Interaction
Within-sequence shuffled	0.6273	0.6337	0.7234	0.7314	+0.0016
Intact property prior	0.6273	0.7069	0.7644	**0.9034**	**+0.0594**

Table 16: Priority-direction comparison. Balanced utility is the arithmetic mean of the P/S/N means.

Train	Sample	P	S	N	Utility	Joint pass	Valid	PPL_{gen}
T^+	S^+	0.9017	0.8983	0.9102	0.9034	0.8462	0.849	2.976
T^+	S^-	0.6957	0.6882	0.7123	0.6987	0.3234	0.853	2.834
T^-	S^+	0.7436	0.7618	0.7663	0.7572	0.5127	0.851	2.918
T^-	S^-	0.5954	0.6187	0.6524	0.6222	0.1437	0.858	2.762

The full operator instead uses the signed mean h_i (Appendix B.1); a positive best-mutant gain can coexist with a negative mean effect. Controls share its quota rule, while realized rewrites depend on ranking and eligibility.

Table 17: Operator controls on the priority-trained backbone. The first row uses readiness alone for commitment; subsequent rows use importance times readiness. PPL_{gen} uses the frozen DPLM on generated sequences.

Operator / remask ranking	Quota	Utility	Joint pass	Valid	Pair diversity	PPL_{gen}
Readiness only / none	0	0.7431	0.4812	0.854	0.7782	2.814
Importance × readiness / none	0	0.8260	0.6478	0.851	0.7684	2.876
Random remask	matched	0.8573	0.7124	0.850	0.7612	2.923
Priority remask (q_i)	matched	0.8917	0.8136	0.849	0.7524	2.958
Improvement remask (d_i)	matched	0.8978	0.8347	0.849	0.7498	2.971
Full `improvable_remask`	default	**0.9034**	**0.8462**	0.849	0.7486	2.976

Random remasking improves utility by 0.0313 over importance–readiness commitment alone; priority-ranked remasking adds 0.0344. The full operator gains 0.0774 in utility and 0.1984 in joint pass over the no-remask control, with Valid lower by 0.002. Its gain over best-mutant remasking is smaller, at 0.0056. Diversity decreases and PPL_{gen} increases with stronger control.

Table 18: Rewrite-quota sensitivity at 64 denoiser NFEs. The labels 1, 2, and 5 denote fixed quotas; Default uses the schedule-dependent quota. Zero denotes the no-remask control. Mean remasks/site counts returns from a revealed residue to the mask state.

Quota setting	Utility	Joint pass	Valid	Remasks/site	NFE
0	0.8260	0.6478	0.851	0.000	64
1	0.8642	0.7236	0.850	0.847	64
2	0.8876	0.7918	0.850	1.692	64
Default	0.9034	0.8462	0.849	2.461	64
5	0.8983	0.8324	0.846	3.814	64

The largest tested fixed quota increases revisits beyond the default but reduces utility by 0.0051 and joint pass by 0.0138 (Table 18). The default therefore balances rewrite opportunities and retained context better within this sweep. The zero-quota control disables all remasking; Appendix B.1 distinguishes this from disabling only the forced override.

B.5 HYPERPARAMETER SENSITIVITY

Table 19: Hyperparameter operating points and qualitative sweep behavior.

Hyperparameter	Symbol	Default	Observed behavior
Priority bias (stochastic)	w	2.0	stronger steering to about 3, but larger values increase PPL cost
Priority fraction (binary)	f	0.4	0.4 is stronger than 0.3 except for concentrated priors
Curriculum maximum (stochastic)	$\rho_{\max}$	0.7	0.7 balances steering, whereas 1.0 overfits guide noise
Uncertainty discount	λ_u	0.5	helpful when guide disagreement is high
Gate temperature	τ_g	1.0	mild effect over $[0.5, 2]$
Readiness exponents	α, β, γ	1,1,1	margin has largest component effect, while stability helps late
Guide interval	M	5	13 refreshes at $M = 5$ versus 64 at $M = 1$; see Appendix B.6.1
Guide exponent	λ_g	1.0	larger values emphasize importance, although sampling-only saturates
Sampling iterations	T	64	sampling-only improves, then plateaus from 128 to 256

Table 19 summarizes the explored settings; its first three rows apply to stochastic/binary alternatives. Appendix B.5.1 quantifies uncertainty, temperature, and guide-interval sensitivity. The default $M = 5$ retains most of the score gain with fewer guide refreshes.

B.5.1 MEASURED PARAMETER SENSITIVITY

Table 20: Measured parameter-sweep summaries. Defaults are $\lambda_u = 0.5$, $\tau_g = 1$, and $M = 5$. RF input counts use nominal length 15 and exclude additional disagreement passes.

Parameter	Value	Utility	Joint pass	Valid	RF inputs
λ_u	0	0.8976	0.8287	0.851	11,154
λ_u	0.5	0.9034	0.8462	0.849	11,154
λ_u	1	0.8947	0.8218	0.847	11,154
τ_g	0.5	0.8987	0.8376	0.848	11,154
τ_g	1	0.9034	0.8462	0.849	11,154
τ_g	2	0.8968	0.8317	0.849	11,154
M	1	0.9167	0.8587	0.849	54,912
M	5	0.9034	0.8462	0.849	11,154
M	10	0.8876	0.8218	0.850	6,006

The $\lambda_u = 0$ utility agrees with the component ablation in Table 8. Utility spans 0.8947–0.9034 over the uncertainty-discount settings and 0.8968–0.9034 over gate temperatures. Relative to $M = 1$, $M = 5$ reduces nominal RF input evaluations by about 79.7%, with utility lower by 0.0133. Increasing the interval to ten reduces queries further with an additional utility decrease of 0.0158.

B.6 COMPUTATIONAL COST

B.6.1 RF AND TARGET-GUIDE QUERY ACCOUNTING

For length L, a complete three-property RF refresh scores the wild type and $19L$ mutants under each guide, giving $3(19L+1)$ sequence–guide evaluations. With $R = \lceil T/M \rceil$ refreshes, the nominal total

is $3R(19L + 1)$. These counts exclude the extra tree-level passes for disagreement, data movement, and batching overhead.

Exact target-conditioned pseudo-perplexity requires L leave-one-out inputs per scored sequence. Scoring the wild type and every mutant therefore costs $RL(19L+1)$ single-input forward equivalents before batching or caching. Reusing the wild-type distribution when the mutated site is masked can reduce redundant work; the nominal count includes these inputs. Each input's cost also depends on target length.

Table 21: Analytical input counts at $L = 15, T = 64$. The final column applies to the exact target-conditioned guide.

M	Refreshes	Denoiser NFE	RF inputs	PepMLM masked inputs
1	64	64	54,912	274,560
5	13	64	11,154	55,770
10	7	64	6,006	30,030

For variable lengths, the expected per-refresh input count is $19\mathbb{E}[L^2] + \mathbb{E}[L]$. Offline target-prior construction and target-specific fine-tuning are additional costs.

Table 22: Target-specific input accounting over the measured peptide-length distributions. Each trajectory has 13 guide refreshes and 64 denoiser NFEs. Forward equivalents count individual masked PepMLM inputs, not batched model invocations. Per-trajectory values are 13 times the rounded per-refresh summaries.

Target	Mean peptide length	Inputs/refresh	Inputs/trajectory
GFAP	15.3	4,527	58,851
TfR	14.8	4,233	55,029
GLP1R	15.7	4,769	61,997
NCAM1	15.1	4,408	57,304
AMHR2	14.9	4,293	55,809
GLAST	16.4	5,210	67,730
RBX1	14.7	4,177	54,301

The seven targets average approximately 58,717 masked inputs per trajectory (Table 22), versus 55,770 at fixed $L = 15$. Latency additionally depends on target length, precision, batching, and input reuse.

B.6.2 Measured comparison profiles

Table 23: Generation profiles on A100-40GB with method-specific batch sizes. Seconds/sequence and peak memory describe these operating points. RF counts are nominal forward equivalents at $L = 15$. Each gradient-guidance round evaluates three property heads and computes reward gradients with respect to detached logits (Appendix A.3); tree-search work counts scorer calls. Costs exclude fine-tuning and target-conditioned PepMLM scoring.

Method	Denoiser NFE	RF inputs	Other scorer work	Batch	s/seq	Peak GB
DPLM backbone	64	0	none	128	0.12	4.2
Gradient guidance	64	0	64 guidance rounds	64	0.38	8.4
PAPL	64	0	planner	128	0.15	5.1
PAPL + matched sampler	64	11,154	none	64	0.87	5.3
PPDIFF property-only	64	11,154	none	64	0.92	5.6
PepTune	~ 512	0	$\sim 3,200$ MCTG calls	32	2.41	12.8

Property-only **PPDIFF** takes 0.92 seconds per sequence at the reported operating point, compared with 0.87 for PAPL with the same guided sampler. Both incur RF-based ISM overhead beyond 64 denoiser steps. Target-conditioned input counts appear separately in Table 22.

C Property Distribution-Shift Statistics

Table 24 quantifies how training-only policies move each complete score distribution relative to the property-neutral backbone. JS measures distributional separation, Wasserstein measures transport distance, and KS tests the largest cumulative-distribution difference.

Table 24: Distribution shifts relative to the backbone start for permeability (P), hemolysis safety (S), and nonfouling (N). The shared layout makes the policy ordering directly comparable across properties.

Property	Policy	JS	Wasserstein	KS	KS p
P	property ordered	0.0714	0.1172	0.268	3.35×10^{-32}
	weighted priority	0.0461	0.0938	0.217	3.77×10^{-21}
	hard top 40%	0.0342	0.0806	0.188	6.03×10^{-16}
	ordered + random mix	0.0298	0.0753	0.176	5.11×10^{-14}
	soft priority 40%	0.0243	0.0672	0.159	1.65×10^{-11}
	random control	0.0005	0.0042	0.033	0.6352
	shuffled-prior control	0.0019	0.0114	0.048	0.1932
S	property ordered	0.0397	0.0856	0.206	4.38×10^{-19}
	weighted priority	0.0221	0.0637	0.155	5.89×10^{-11}
	hard top 40%	0.0168	0.0548	0.136	1.58×10^{-8}
	ordered + random mix	0.0128	0.0479	0.119	1.26×10^{-6}
	soft priority 40%	0.0101	0.0421	0.106	2.40×10^{-5}
	random control	0.0004	0.0048	0.030	0.7473
	shuffled-prior control	0.0014	0.0097	0.043	0.3048
N	property ordered	0.0542	0.0716	0.247	2.36×10^{-27}
	weighted priority	0.0254	0.0488	0.173	1.48×10^{-13}
	hard top 40%	0.0126	0.0341	0.123	4.74×10^{-7}
	ordered + random mix	0.0088	0.0283	0.103	4.52×10^{-5}
	soft priority 40%	0.0065	0.0244	0.090	5.65×10^{-4}
	random control	0.0003	0.0032	0.027	0.8494
	shuffled-prior control	0.0010	0.0076	0.037	0.4889

All three properties reproduce the training-policy ordering in Table 7. Property ordering produces the largest shift by all three distances, while random continuation and prior shuffling remain close to the backbone and have nonsignificant KS tests. The distributional comparisons agree with the mean-score improvements.

D Additional Generation Results

D.1 Requested objective weights

Table 25 compares seven requested property-weight settings. Single-objective settings favor their selected property, while equal weights yield the largest three-property joint pass and mean minimum. Across these complete generation configurations, Pearson correlations between requested weights and corresponding property means are 0.8863 (P), 0.8904 (S), and 0.8786 (N).

Table 25: Property-weight control. All three properties are evaluated at every weight setting. Mean min. is the average per-sequence minimum $\mathbb{E}[\min(P, S, N)]$.

w_P	w_S	w_N	P	S	N	Mean min.	Joint pass	Valid
1	0	0	0.9387	0.7263	0.7412	0.6847	0.3847	0.847
0	1	0	0.7184	0.9352	0.7523	0.6762	0.3624	0.846
0	0	1	0.7328	0.7417	0.9416	0.6914	0.3781	0.848
1/2	1/2	0	0.9143	0.9058	0.7816	0.7462	0.5837	0.848
1/2	0	1/2	0.9168	0.7694	0.9214	0.7327	0.5614	0.847
0	1/2	1/2	0.7537	0.9087	0.9256	0.7218	0.5428	0.848
1/3	1/3	1/3	0.9017	0.8983	0.9102	0.8681	0.8462	0.849

Table 26: Equal-weight objective over permeability, safety, and nonfouling. Square brackets are 95% confidence intervals.

Training policy	Train only	`weighted_T_V`	Improvable remask
property ordered	0.7069	0.8260 [0.8190, 0.8330]	**0.9034 [0.8970, 0.9098]**
weighted priority	0.6874	0.8174 [0.8102, 0.8246]	0.8933 [0.8865, 0.9001]
hard top 40%	0.6768	0.8123 [0.8050, 0.8196]	0.8891 [0.8821, 0.8961]

D.2 Joint objectives across backbones

Guided sampling improves every training policy, and the confidence intervals remain above the corresponding training-only point estimates. The strongest policy before guidance remains strongest afterward.

D.3 Sampling operator averages

Table 27: Sampling operators averaged over the top three priority-trained backbones.

Objective	Train only	`weighted_T_V`	Improvable remask
Permeability	0.6840	0.8091	0.8923
Safety	0.6827	0.8183	0.8884
Nonfouling	0.7045	0.8284	0.9052

Table 27 averages across the three training policies to reduce dependence on a single backbone variant. `improvable_remask` is consistently stronger than `weighted_T_V`, while both substantially improve on direct sampling from the trained checkpoint.

D.4 Sequence quality

Table 28: Generation quality and residue-level masked-diffusion validation perplexity. Valid is the protocol-level fraction, Unique is computed among protocol-valid outputs, and NN identity and diversity use scoreable outputs. Lower median nearest-neighbor identity indicates greater novelty.

Model	Valid ↑	Unique ↑	PPL ↓	NN-id$_{50}$ ↓	Pair diversity ↑
Backbone start	0.862	0.908	2.451	0.4218	0.7823
Random-mask finetune	0.859	0.912	2.461	0.4235	0.7801
Weighted priority	0.856	0.907	2.420	0.4387	0.7712
Property ordered	0.855	0.901	2.421	0.4514	0.7634
Hard top 40%	0.853	0.904	2.469	0.4406	0.7746
PPDIFF full	0.849	0.898	2.421	0.4672	0.7486

Property control produces modest but measurable concentration. Relative to the backbone, validity decreases from 0.862 to 0.849, pair diversity from 0.7823 to 0.7486, and median nearest-neighbor identity rises from 0.4218 to 0.4672. Checkpoint validation PPL remains similar (2.421 versus 2.451), while generated-sequence PPL$_{\text{gen}}$ increases from 2.683 to 2.976 (Table 40), consistent with a control–diversity tradeoff.

Table 29: Best validation perplexity by property. Permeability and safety differ from the 2.451 start value by about 1.3%, while nonfouling differs by 0.2%.

Property	Best PPL	Property-best policy PPL
Permeability	2.420	2.421
Safety	2.418	2.419
Nonfouling	2.447	2.447

The property-best policies also remain close to the best validation PPL within each property (Table 29).

E LEAKAGE AND EVALUATOR AUDITS

We quantify sequence overlap, evaluator reliability, and coverage of generated populations within evaluator applicability domains.

E.1 PAIRWISE LEAKAGE AUDIT

Table 30: Pairwise overlap fractions among the guide data, evaluator data, and **PPDIFF** generations.

Pair	Exact dup.	Cluster@90%	Cluster@60%	Cluster@30%	Enriched 3-mer
Guide train vs. ESM-150M-F test	0	0	0.0214	0.1227	0.0847
Guide train vs. FP test	0	0	0.0184	0.1047	0.0723
Guide train vs. PeptideBERT test	0	0	0.0317	0.1583	0.0973
PPDIFF gen. vs. guide train	0	0.0123	0.0847	0.2337	0.1417
PPDIFF gen. vs. evaluator train	0	0.0081	0.0634	0.1873	0.1176

Table 30 finds no exact guide-to-evaluator duplicates and no overlap at 90% identity. Overlap grows at looser thresholds, as expected for short peptides, which motivates the hybrid split and cross-representation evaluators.

E.2 GUIDE–EVALUATOR TRAINING OVERLAP

Table 31: Guide-training versus evaluator-training overlap fractions, supplementing the evaluator-test audit in Table 30.

RF guide train vs.	Exact	90%	60%	30%	Enriched 3-mer
ESM train	0	0	0.0847	0.2536	0.1524
FP train	0	0	0.0763	0.2387	0.1387
PeptideBERT train	0	0	0.0942	0.2716	0.1628

Although exact and 90%-identity overlaps are zero, overlap persists at lower identity thresholds and at the motif level. Cross-evaluator agreement may therefore retain shared source-label and sequence biases.

E.3 EVALUATOR RELIABILITY AND APPLICABILITY DOMAIN

Table 32: Held-out evaluator reliability, macro-averaged across the three property heads. ECE denotes expected calibration error.

Evaluator	AUC	PR-AUC	ECE / slope	Split	Guide-data relation
ESM-150M-F	0.8347	0.8116	0.0423 / 0.968	homology/ hybrid	cluster-disjoint
PepBenchmark FP	0.8023	0.7783	0.0513 / 0.937	homology/ hybrid	feature/provenance distinct
PeptideBERT (3 heads)	0.7826	0.7513	0.0634 / 0.913	custom cluster split	architecture distinct

Table 32 shows that all evaluators have useful discrimination and calibration. ESM-150M-F is strongest and therefore serves as the primary evaluator, while the fingerprint and PeptideBERT models test whether the ranking transfers across representations.

The applicability-domain audit shows a slight composition shift for **PPDIFF** and similar length coverage.

F GUIDE DEPENDENCE AND OPTIMIZATION STRESS TESTS

This section expands the guide-degradation, compute, and no-denoiser controls summarized in Section 5.3, using the protocol in Appendix A.

Table 33: Applicability-domain audit. NN columns are maximum sequence identity to the corresponding training set. Length OOD is outside the interval between the training 1st and 99th percentiles. Composition OOD is above the training 99th percentile of amino-acid-composition Mahalanobis distance. OOD entries are percentages.

Method	NN guide	NN eval.	Length OOD (%)	Comp. OOD (%)	Uncertainty
Backbone start	0.3816	0.4117	3.23	7.43	0.1237
PPDIFF full	0.4013	0.4347	3.84	11.37	0.1376

F.1 GUIDE RELIABILITY AND CONTROLLED DEGRADATION

Table 34: Held-out reliability of the RF guides. ISM ρ is the mean within-sequence Spearman correlation between RF-based ISM mutation effects and the separate evaluator ensemble. Top-20% is their position-set overlap. The final row is the macro mean.

Property	AUROC	PR-AUC	ECE	MCC	ISM ρ	Top-20%
Permeability	0.8126	0.7938	0.0562	0.4873	0.4217	0.4386
Safety	0.8241	0.8067	0.0498	0.5016	0.4479	0.4563
Nonfouling	0.7863	0.7619	0.0617	0.4384	0.3684	0.3959
Macro mean	0.8077	0.7875	0.0559	0.4758	0.4127	0.4303

Table 35: Dependence on guide quality. ESM utility is mean $\pm$ SD over five independent fine-tuning seeds, each decoded with three generation seeds. FP averages three property-score means; PB averages three marginal pass fractions, as in Table 3. All other optimization and generation budgets are fixed.

Guide condition	AUROC	ISM ρ	ESM utility	FP	PB	Valid
100% guide data	0.8077	0.4127	0.9009 ± 0.0172	0.8462	0.8494	0.849
50% guide data	0.7894	0.3719	0.8836 ± 0.0185	0.8281	0.8317	0.851
25% guide data	0.7648	0.3186	0.8512 ± 0.0207	0.7963	0.7991	0.853
10% guide data	0.7216	0.2268	0.8028 ± 0.0229	0.7427	0.7374	0.855
40% label noise	0.6692	0.1425	0.7691 ± 0.0238	0.7084	0.6996	0.857
Label-shuffled guide	0.5013	0.0021	0.6358 ± 0.0221	0.5809	0.5618	0.861
Within-sequence priority shuffle	0.8077	0.0028	0.7330 ± 0.0246	0.6761	0.6658	0.858

Downstream utility decreases with guide-data reduction and label corruption (Tables 34–35). Within-sequence shuffling preserves sequence AUROC at 0.8077 but reduces local rank agreement to 0.0028 and dual-end utility to 0.7330 (Figure 3), linking performance to informative local rankings. The full-data utility 0.9009 is the independent-fine-tuning-seed mean; headline tables use the seed-42 estimate 0.9034.

F.2 SAMPLING-ONLY COMPUTE AND GUIDANCE-STRENGTH FRONTIER

Table 36: ESM balanced utility for sampling-only control over denoising steps T and guide exponent λ_g. The full method at its fixed default $T = 64, \lambda_g = 1$ reaches 0.9034.

T	$\lambda_g = 0.5$	$\lambda_g = 1$	$\lambda_g = 2$	$\lambda_g = 4$
32	0.7168	0.7331	0.7417	0.7463
64	0.7441	0.7644	0.7796	0.7872
128	0.7698	0.7932	0.8163	0.8207
256	0.7836	0.8087	**0.8214**	0.8189

Sampling-only utility rises with T and λ_g and plateaus at 0.8214, 0.0820 below the default full method despite four times as many denoiser steps (Table 36). Fingerprint and PeptideBERT performance peaks no later than this ESM frontier. At $T = 256, \lambda_g = 4$, ESM utility declines slightly, Valid falls to 0.764, and composition OOD reaches 26.73% (Tables 37–38). The best sampling-only ESM point has valid-and-joint yield 0.5118, compared with 0.7248 for the full method. Within this grid, additional inference compute does not recover the performance of priority-aware training.

Table 37: Cross-evaluator sampling-only frontier. ESM entries are mean ± SD under matched seeds. Joint pass uses validation-fixed thresholds; FP averages property-score means and PB averages marginal pass fractions. The full method is the fixed seed-42 point reference.

Condition	ESM utility	Joint pass	FP	PB
Sampling only, $T = 64$, $\lambda_g = 1$	0.7644 ± 0.0114	0.5364	0.7023	0.6837
Sampling only, $T = 128$, $\lambda_g = 2$	0.8163 ± 0.0138	0.6381	0.7476	0.7318
Sampling only, $T = 128$, $\lambda_g = 4$	0.8207 ± 0.0146	0.6448	0.7432	0.7226
Sampling only, $T = 256$, $\lambda_g = 2$	0.8214 ± 0.0151	0.6427	0.7501	0.7324
Sampling only, $T = 256$, $\lambda_g = 4$	0.8189 ± 0.0178	0.6352	0.7358	0.7107
PPDIFF full, $T = 64$, $\lambda_g = 1$	**0.9034**	**0.8462**	**0.8462**	**0.8494**

Table 38: Quality guardrails for the sampling-only frontier. Valid-and-joint is measured over all raw generations. Composition OOD is relative to the training-set 99th-percentile threshold.

Condition	Valid	Unique	Pair diversity	Comp. OOD (%)	Valid-and-joint
Sampling only, $T = 64$, $\lambda_g = 1$	0.852	0.902	0.7659	9.18	0.4619
Sampling only, $T = 128$, $\lambda_g = 2$	0.827	0.893	0.7412	14.72	0.5228
Sampling only, $T = 128$, $\lambda_g = 4$	0.801	0.879	0.7084	20.41	0.5076
Sampling only, $T = 256$, $\lambda_g = 2$	0.806	0.884	0.7215	18.36	0.5118
Sampling only, $T = 256$, $\lambda_g = 4$	0.764	0.851	0.6748	26.73	0.4659
PPDIFF full, $T = 64$, $\lambda_g = 1$	**0.849**	**0.898**	0.7486	11.37	**0.7248**

F.3 NO-DENOISER CONTROLS

Table 39: Transfer from the RF objective to three separate evaluators. ESM and FP average property-score means; PB averages marginal pass fractions. RF-only greedy coordinate ascent is allowed to maximize the proxy under the matched RF-query budget.

Method	RF objective	ESM	FP	PB
Uniform-AA random	0.5041	0.5118	0.4937	0.4815
Empirical-unigram random	0.5487	0.5564	0.5268	0.5089
Random + RF top-K	0.9186	0.7247	0.6512	0.6284
RF-only greedy coordinate ascent	**0.9786**	0.7482	0.6541	0.6217
DPLM backbone start	0.6154	0.6273	0.5742	0.5531
DPLM + RF reranking (64 candidates)	0.8467	0.7426	0.6808	0.6621
PPDIFF sampling only	0.8729	0.7644	0.7023	0.6837
PPDIFF full	0.9348	**0.9034**	**0.8462**	**0.8494**

Table 40: Quality and effective yield for no-denoiser controls. Valid × ESM is yield-adjusted balanced utility; PPL_{gen} is frozen-DPLM NELBO perplexity on generated sequences.

Method	Joint	Valid	Valid-and-joint	Valid × ESM	PPL_{gen}	Pair diversity	Comp. OOD (%)
Uniform-AA random	0.0718	0.996	0.0709	0.5098	8.731	0.9472	47.63
Empirical-unigram random	0.1027	0.982	0.0998	0.5464	6.482	0.8826	23.41
Random + RF top-K	0.3942	0.641	0.2527	0.4645	9.284	0.7025	55.82
RF-only greedy coordinate ascent	0.4348	0.287	0.1248	0.2147	11.327	0.4126	76.48
DPLM backbone start	0.1543	0.862	0.1391	0.5407	2.683	0.7823	7.43
DPLM + RF reranking (64 candidates)	0.4681	0.858	0.4016	0.6371	2.941	0.7418	10.83
PPDIFF sampling only	0.5364	0.852	0.4619	0.6512	2.865	0.7659	9.18
PPDIFF full	**0.8462**	0.849	**0.7248**	**0.7669**	2.976	0.7486	11.37

RF-only greedy search obtains the highest proxy score (Table 39), but transfers poorly: validity is 0.287, valid-and-joint yield 0.1248, PPL_{gen} 11.327, and composition OOD 76.48% (Tables 40–41). These changes indicate proxy exploitation. DPLM reranking improves naturalness but remains below trajectory guidance. The denoiser supplies peptide-like proposals and readiness estimates that constrain where and when the proxy is optimized.

G TARGET-SPECIFIC GENERATION AND BINDING SCREEN

Generation-side binding control. For GFAP, TfR, GLP1R, NCAM1, AMHR2, GLAST, and RBX1, the full canonical human target sequence conditions the frozen PepMLM-650M guide

Table 41: Inference query counts per returned sequence at nominal length $L = 15$. RF calls are input-level forward equivalents across three guides, with 13 refreshes for guided sampling. Counts exclude tree-level disagreement passes, fine-tuning, and prior precomputation; measured runtime profiles appear in Appendix B.6.2.

Method	Denoiser NFE	RF calls	Accounting note
Uniform-AA random	0	0	no model or guide
Empirical-unigram random	0	0	no model or guide
Random + RF top-K	0	11,154	query-matched rejection/reranking
RF-only greedy coordinate ascent	0	11,115	13 full coordinate sweeps at nominal $L = 15$
DPLM backbone start	64	0	one complete generation
DPLM + RF reranking (64 candidates)	4,096	192	64 generations plus three-guide reranking
PPDIFF sampling only	64	11,154	13 RF-based ISM refreshes at nominal $L = 15$
PPDIFF full	64	11,154	inference only with amortized fine-tuning

(Section 3.3). Its ISM priority is fused with permeability, safety, and nonfouling using equal rank-space weights. Each target has a separate fine-tuning-and-sampling run from the common initialization.

Frozen-pool docking screen. For each target, we freeze the top-20 candidates by the PepMLM/P/S/N generation objective before rigid multi-conformer Vina screening. Each peptide uses five ETKDGv3/MMFF conformers and is paired with composition-preserving scrambled and charge-pattern-shuffled controls. To account for ligand size and receptor preparation, we emphasize ligand efficiency (LE), the docking score divided by the uncapped linear peptide's heavy-atom count, and within-target candidate-minus-control differences. The known TfR binder T7 (HAIYPRH) gives -10.18 kcal/mol and LE -0.159 as a calibration point under the original protocol.

Across the seven leads in Table 4, property means are 0.8313, 0.8498, and 0.8377, and the mean per-sequence minimum is 0.8248. In the frozen top-20 pools, candidate LE is better than paired-control LE on six targets after Benjamini–Hochberg correction ($q < 0.05$).

G.1 Target specification and frozen-pool audit

PepMLM receives the canonical human UniProt sequence (The UniProt Consortium, 2025), whereas Vina receives only the deposited Protein Data Bank structural chain (Berman et al., 2000). Table 42 joins these identifiers and structural scopes directly to the complete frozen-pool comparison.

Table 42: Target specification and frozen top-20 pool audit. Box dimensions are in Å; LE is in kcal/mol per heavy atom (lower is better). ΔLE is candidate minus paired-decoy mean. Each control averages one composition-preserving scramble and one charge-pattern shuffle. Sites and pools were fixed before docking.

Target	PepMLM target	PDB/chain	Structural scope / box	Cand. LE	Decoy LE	ΔLE	Wilcoxon p	BH q
GFAP	P14136	6A9P/A	coiled-coil center / 46^3	-0.05468	-0.04213	-0.01255	7.6×10^{-3}	0.0106
TfR	P02786	3KAS/A	viral-GP1 interface / 44^3	-0.07826	-0.05873	-0.01953	2.8×10^{-3}	0.0098
GLP1R	P43220	3C5T/A	ECD peptide interface / 42^3	-0.08214	-0.05783	-0.02431	1.1×10^{-3}	0.0077
NCAM1	P13591	2HAZ/A	isolated domain center / 44^3	-0.07284	-0.05517	-0.01767	4.7×10^{-3}	0.0106
AMHR2	Q16671	7L0J/B	AMH interface / 42^3	-0.06874	-0.05128	-0.01746	6.3×10^{-3}	0.0106
GLAST	P43003	5LM4/A	6Z6 site, one state / 42^3	-0.06813	-0.05627	-0.01186	3.2×10^{-2}	0.0373
RBX1	P62877	1LDJ/B	Cullin interface / $46 \times 46 \times 42$	-0.05216	-0.04793	-0.00423	2.3×10^{-1}	0.2300

GFAP and NCAM1 have no peptide-defined site. The TfR box covers the co-crystallized viral-GP1 interface, and GLP1R contains only the extracellular domain. GLAST represents one transporter state; deposited RBX1 ASN41 is excluded because of nonphysical atom overlaps. RBX1 remains nonsignificant ($q = 0.2300$). Five rigid conformers provide sparse coverage of flexible peptides. The added three-seed, four-decoy setting retains the same six significant targets, with rank-stability summaries of 0.824–0.913 (Appendix G.3).

For the reproducible pose reconstruction, PDBFixer 1.12 (OpenMM contributors, 2026) completes heavy atoms but does not rebuild unresolved loops. Heterogens are removed except the three

RBX1 zinc ions. Meeko 0.7.1 (Forli Lab, 2026) prepares receptors at nominal pH 7.4. Five RDKit ETKDGv3/MMFF conformers per uncapped linear peptide are docked rigidly with Vina 1.2.7 using exhaustiveness 16, nine modes, a 6-kcal/mol energy range, and a fixed base seed (Trott & Olson, 2010; Eberhardt et al., 2021).

G.2 Lead provenance and reconstruction audit

Table 43 links the seven leads back to the sequences and P/S/N scores in Table 4. Heavy-atom count (HAC) is the denominator of LE. Net charge counts R/K as $+1$ and D/E as -1, with histidine approximately neutral at pH 7.4. Hydrophobic moment uses the Eisenberg scale at 100° per residue. GRAVY uses the Kyte and Doolittle scale, while cationic fraction counts R/K/H.

Table 43: Physicochemical properties and protocol-sensitivity redocking audit for the seven displayed leads. Original and reconstructed columns are Vina scores in kcal/mol. The value Δ is reconstruction minus original, so positive values are less favorable.

Target	HAC	Charge	Hydr. moment	GRAVY	Cat. frac.	Original	Reconstructed	Δ
GFAP	167	+2	0.5591	-1.356	0.3125	-10.540	-9.644	+0.896
TfR	148	+2	0.2742	-1.607	0.3333	-13.950	-11.085	+2.865
GLP1R	149	0	0.3650	-0.9313	0.2500	-14.820	-11.871	+2.949
NCAM1	144	0	0.3897	-1.893	0.2000	-12.720	-11.508	+1.212
AMHR2	150	+2	0.1374	-1.273	0.3333	-12.470	-11.982	+0.488
GLAST	166	0	0.2491	-0.7167	0.1111	-13.980	-11.885	+2.095
RBX1	158	-4	0.3478	-1.480	0.0667	-10.500	-11.120	-0.620

The leads span charges from -4 to $+2$ and varied hydropathy profiles. Reconstruction shifts reach $+2.949$ kcal/mol, demonstrating preparation sensitivity and motivating within-protocol paired comparisons.

G.3 Additional docking settings

Mean within-target ΔLE is -0.01537 kcal/mol per heavy atom for the original setting (five conformers, one seed, two decoys), -0.01684 for ten conformers with four decoys, and -0.01512 for five conformers across three seeds with four decoys. Both added settings also increase decoy count. All settings retain seven negative mean effects and six targets with BH $q < 0.05$. In the multi-seed setting (Table 44), these six targets have rank stability 0.824–0.913; RBX1 remains nonsignificant. Seed SD measures variability within the same candidate pools. Preparation sensitivity is assessed separately in Appendix G.1.

Table 44: Multi-seed docking with five conformers, three seeds, and four decoys per candidate. ΔLE is candidate-minus-decoy LE in kcal/mol per heavy atom; SD is across seeds. ρ_{rank} summarizes cross-seed Spearman rank stability. The reported p- and BH q-values apply to this setting, separately from Table 42.

Target	Cand. LE	Decoy LE	ΔLE	Seed SD	p	BH q	ρ_{rank}
GFAP	-0.05454	-0.04226	-0.01228	0.00387	0.0081	0.0113	0.847
TfR	-0.07814	-0.05890	-0.01924	0.00324	0.0031	0.0109	0.891
GLP1R	-0.08203	-0.05795	-0.02408	0.00298	0.0012	0.0084	0.913
NCAM1	-0.07271	-0.05529	-0.01742	0.00356	0.0052	0.0113	0.872
AMHR2	-0.06858	-0.05139	-0.01719	0.00368	0.0068	0.0113	0.863
GLAST	-0.06798	-0.05635	-0.01163	0.00394	0.0360	0.0420	0.824
RBX1	-0.05202	-0.04802	-0.00400	0.00418	0.2400	0.2400	0.712

H Broader Impact

Controllable peptide generation can support therapeutic discovery, delivery, and biomaterials, but it can also lower the cost of designing harmful bioactive sequences. Safety-aware objectives do not substitute for access control or experimental review.