REACHDIFF: FUTURE-FEASIBILITY-GUIDED SEARCH FOR CONSTRAINED CYCLIC PEPTIDE DESIGN

Anonymous authors
Paper under double-blind review

ABSTRACT

Designing cyclic peptides requires several coupled requirements to be satisfied simultaneously, including chemical validity, cyclization-topology realization, and closure geometry. Their intersection defines a substantially narrower design space than any individual requirement, making efficient exploration particularly important. Most generative pipelines nevertheless spend the majority of sampling computation before determining whether a completed molecule satisfies the joint feasibility criteria. We introduce **ReachDiff**, a masked-diffusion approach that predicts future feasibility, defined as the probability that a partial molecular state will satisfy the benchmark joint-feasibility objective when generation continues with the frozen base model. A lightweight predictor scores competing partial states once during generation. The highest-scoring states receive the remaining continuation budget, after which standard denoising resumes without further guidance. Across head-to-tail, disulfide, and isopeptide-lactam cyclization, ReachDiff improves joint feasibility from 59.61% to 77.59% under an equal terminal-evaluation budget, and the advantage is retained under matched recorded model work. Future feasibility is predictable early in denoising, and controlled analyses attribute the gain to state-dependent ranking and global competition. ReachDiff also collects verifier-positive trajectories more efficiently, which can be reused for denoising alignment to improve search-free generation. Together, these results support future-feasibility-guided allocation as an effective use of finite inference compute.

1 INTRODUCTION

Cyclic peptides occupy an attractive region of molecular design space between small molecules and larger biologics. Macrocyclization can restrict conformational freedom, improve stability, and pre-organize binding-competent conformations (Zorzi et al., 2017; Ji et al., 2024; Dougherty et al., 2019; Bechtler & Lamers, 2021). Computational design must jointly satisfy chemical validity, ring topology, and compatible local closure geometry. Because these requirements are coupled, their intersection defines a substantially narrower design space than any individual constraint.

Recent peptide and cyclic-peptide generators improve molecular representation, structural conditioning, geometric constraints, property guidance, and reward-guided search (Kong et al., 2024; Li et al., 2024a; Jiang et al., 2025; Zhou et al., 2025; Zhao et al., 2026; Zhang et al., 2026a). Yet many pipelines determine whether the joint feasibility criteria are satisfied only after generation is complete. If unfinished trajectories already differ in their likelihood of eventual success, continuing them with equal effort can spend substantial computation on states that are unlikely to produce feasible products.

The key premise of this work is that eventual design success may already be predictable from an unfinished diffusion state, early enough to guide the remaining computation. Under the frozen denoising model, we define future feasibility as the probability that a partial molecular state will satisfy the benchmark joint-feasibility objective after normal continuation. This provides a value-like signal for deciding which trajectories should receive the remaining generation budget.

We introduce ReachDiff to estimate future feasibility and use it for computation allocation during masked diffusion. As illustrated in Figure 1, a lightweight predictor estimates future feasibility from the hidden representation of each partial molecular state. Competing continuations share a fixed completion budget, which is allocated globally to the highest-scoring states; the selected states then complete generation under the unchanged frozen model. This separates inference into two

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.

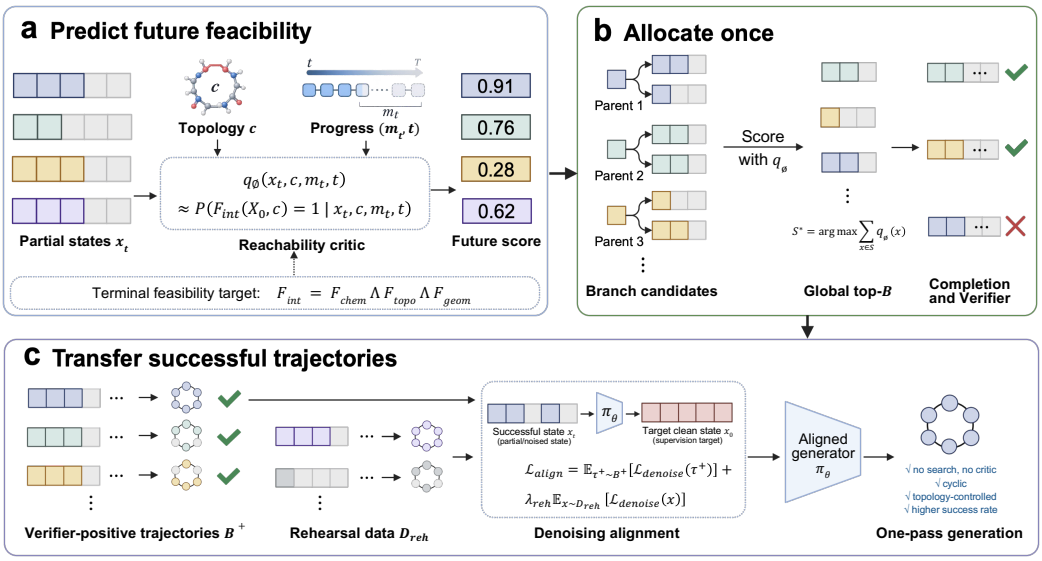


Figure 1: **Overview of ReachDiff. (a)** A future-feasibility predictor estimates the probability that a partial molecular state will eventually satisfy the joint cyclic-peptide design objective. **(b)** Candidate continuations are scored once, and the global top-B states receive the remaining completion budget before frozen denoising resumes. **(c)** As an optional downstream use, successful trajectories can be reused with rehearsal data for denoising alignment, producing a standalone generator that requires no search or predictor at inference.

components: estimating the likelihood of successful completion and allocating finite computation accordingly. Successful search trajectories can further be reused as standard denoising supervision, yielding a standalone generator that no longer requires search at inference time.

Across three cyclization chemistries, ReachDiff consistently improves joint-feasible discovery under both equal terminal-evaluation and matched recorded-work budgets. Mechanistic controls show that the gain is associated with early state-dependent predictability and global allocation rather than branching alone, while a frozen post-hoc structural screen supports transfer beyond the primary verifier. Successful search trajectories can additionally be reused for search-free generator alignment.

Our contributions are threefold.

- **A partial-state view of multi-constraint cyclic-peptide design.** We define future feasibility as the probability that an unfinished peptide will satisfy the benchmark chemical-validity, topology-class, and closure-geometry criteria under the frozen generator, linking terminal success to intermediate molecular states.
- **A minimal inference procedure for finite-budget allocation.** A partial-state predictor and a single global selection step allocate a fixed continuation budget to promising trajectories. Successful trajectories can be reused for standard denoising training without search at deployment.
- **Improved joint-feasible discovery with mechanistic evidence.** Across three cyclization chemistries, ReachDiff improves joint-feasible yield and unique feasible discovery. Temporal, input, and selection controls support the role of early state-dependent predictability and global ranking beyond branching alone.

2 Related Work

Cyclic-peptide generation. Recent work has studied cyclic-peptide generation through geometric diffusion and macrocycle design (Jiang et al., 2025; Rettie et al., 2025), atom–bond diffusion (Zhou et al., 2025), conformation and ensemble modeling (Wang et al., 2026; Secor et al., 2026), masked discrete diffusion (Tang et al., 2025a), and property- or reward-guided generation (Zhao et al., 2026;

Zhang et al., 2026a). These approaches primarily improve molecular representation, structural constraints, or generation objectives. In contrast, we keep the generator fixed and ask whether an unfinished molecular state can predict eventual success well enough to guide the remaining computation.

Search and value guidance during diffusion. Masked diffusion provides the discrete generative backbone used in this work (Sahoo et al., 2024; Lee et al., 2025b). Recent work improves diffusion-time inference in several ways: test-time search and trajectory expansion allocate additional computation to promising branches (Tang et al., 2025a;b; Ma et al., 2025; Ta et al., 2026; Zhang et al., 2026b); value- or posterior-based methods score intermediate states or denoising choices to bias future transitions (Li et al., 2024b; Kim et al., 2025; Rector-Brooks et al., 2025; Lee et al., 2026); and importance weighting, constrained guidance, and reward-based steering repeatedly reshape the sampling distribution toward desired outcomes (Ou et al., 2026; Schiff et al., 2025; Cardei et al., 2025; Wu et al., 2023; Singhal et al., 2025; Lee et al., 2025a). We instead ask when terminal constrained-design success becomes predictable from an unfinished state, and whether that signal can be used for a single global allocation of the remaining generation budget.

3 METHOD

3.1 MULTI-CONSTRAINT CYCLIC-PEPTIDE DESIGN

We consider de novo cyclic-peptide design with a topology-conditioned masked-diffusion generator. A generation condition is

$$c = (k, \ell),$$

where k denotes the requested cyclization topology and ℓ denotes the token-canvas length. The canvas length ℓ conditions generation, while the benchmark feasibility endpoint evaluates chemical validity, cyclization-topology-class realization, and closure geometry. The evaluated topologies are head-to-tail, disulfide, and isopeptide-lactam cyclization.

Let π_{θ_0} denote the frozen cyclic-adapted masked-diffusion generator. Starting from an all-MASK canvas, π_{θ_0} progressively reveals molecular tokens over a fixed reverse-diffusion schedule. We denote an intermediate molecular state by x_t, its realized content-mask fraction by $m_t \in [0, 1]$, and the completed product by X_0.

The frozen benchmark endpoint is defined as a conjunction of three checks: the completed molecule must be chemically valid, realize topology class k, and pass the benchmark closure-geometry screen. We write this operational endpoint as

$$F_{\text{joint}}(X_0, c) = F_{\text{chem}}(X_0) \wedge F_{\text{topo}}(X_0, k) \wedge F_{\text{geom}}(X_0, k). \tag{1}$$

F_{chem} checks molecular parsing, connectivity, and valence constraints. In the frozen benchmark labels, F_{topo} denotes topology-class realization: whether the molecular graph realizes the requested head-to-tail, disulfide, or isopeptide-lactam closure class. F_{geom} applies the fixed lightweight closure-geometry screen used by the benchmark.

We use joint feasibility for the event in Eq. 1. It is the historical frozen computational endpoint used to supervise the predictor and compare search controllers, not a claim of experimental synthesizability or therapeutic utility. Section 4.4 therefore reports two frozen robustness checks that are outside predictor training and search: a strict-topology replay enforcing exact residue count, terminal/cap annotations, and positional anchors, and a post-hoc molecular-mechanics consistency screen.

3.2 PREDICTING FUTURE FEASIBILITY FROM PARTIAL STATES

Completing every partial trajectory is wasteful if some intermediate states are already much less likely to satisfy Eq. 1. We therefore ask how likely the current unfinished molecule is to become jointly feasible when the frozen generator continues normally. For the base generator π_{θ_0}, this probability is

$$q^{\pi_{\theta_0}}(x_t, c, m_t, t) = \Pr_{\pi_{\theta_0}}\left[F_{\text{joint}}(X_0, c) = 1 \mid X_t = x_t, c, m_t, t\right]. \tag{2}$$

We call this quantity future feasibility. In value-function terminology, it is the terminal-success value of the current molecular state under the frozen denoising policy. The policy qualifier is important

because the probability refers to normal continuation with the current generator rather than success under an arbitrary rescue strategy.

A lightweight predictor approximates this quantity,

$$q_\phi(x_t, c, m_t, t) \approx q^{\pi_{\theta_0}}(x_t, c, m_t, t). \tag{3}$$

Completed rollouts of the frozen generator supply the supervision. Each partial state inherits the binary terminal outcome

$$y = F_{\text{joint}}(X_0, c) \in \{0, 1\}. \tag{4}$$

Each outcome is one Bernoulli observation of terminal success, while its conditional mean given the partial state is the future-feasibility probability in Eq. 2. The predictor is trained with binary cross-entropy,

$$\mathcal{L}_{\text{reach}} = -\mathbb{E}\left[y \log q_\phi(x_t, c, m_t, t) + (1 - y) \log\left(1 - q_\phi(x_t, c, m_t, t)\right)\right]. \tag{5}$$

The frozen generator also supplies the state representation. We mean-pool its final-layer hidden states over peptide-content positions and concatenate the pooled representation with embeddings of the requested topology, diffusion step, and realized mask fraction. A lightweight prediction head maps these features to a scalar probability. The generator remains frozen during predictor training. The target in Eq. 2 is a probability, but the downstream allocation step depends primarily on the ordering of competing states. We therefore evaluate discrimination and ranking against fresh continuation outcomes and do not rely on absolute probability calibration.

3.3 Allocating a fixed continuation budget

At inference time, only a fixed number B of partial trajectories can receive the full remaining denoising computation. Once future feasibility can be estimated, the allocation problem becomes choosing which candidate states should consume this continuation budget.

We initialize B independent all-MASK replicas and advance each with the frozen generator to a pre-specified warmup point. Each parent then contributes one or two continuations according to a branching rule fixed before any future-feasibility score is observed, producing a candidate cohort $\mathcal{C}$ with

$$B \le |\mathcal{C}| \le 2B.$$

The candidates advance to a pre-specified scoring point and receive one future-feasibility score each using Eq. 3.

Given a completion budget of B, the selected set is

$$S^* = \operatorname*{arg\,max}_{S \subseteq \mathcal{C},\, |S| = B} \sum_{x \in S} q_\phi(x, c, m_t, t). \tag{6}$$

With fixed cardinality, Eq. 6 reduces to keeping the global top-B candidates. The objective maximizes the predictor-estimated expected number of successful base-policy continuations under the available completion budget. It does not require an independence assumption across sibling trajectories because expectation is additive.

Global allocation differs from per-parent selection because continuation slots are shared across the complete cohort. Two strong continuations from the same parent can both survive, while a weak lineage can receive no slot. After this single decision, the predictor is not queried again. The selected states complete generation with the unchanged frozen sampler, and the terminal verifier is applied only after completion. Exact branching and scoring thresholds are reported in Appendix C.3.

3.4 Optional reuse of successful trajectories for generator alignment

Search also produces successful denoising trajectories that can be reused for training. On an alignment split disjoint from held-out evaluation, we rerun the frozen search procedure, retain products satisfying Eq. 1, and store the actual denoising lineage that generated each success. Let $\mathcal{B}_+$ denote this positive-trajectory buffer.

We update the existing cyclic-adaptation parameter subset with standard denoising supervision on $\mathcal{B}_+$ and rehearsal from the cyclic training split,

$$\mathcal{L}_{\text{align}} = \mathbb{E}_{\tau^+ \sim \mathcal{B}_+}[\mathcal{L}_{\text{denoise}}(\tau^+)] + \lambda_{\text{reh}} \mathbb{E}_{x \sim \mathcal{D}_{\text{reh}}}[\mathcal{L}_{\text{denoise}}(x)], \qquad \lambda_{\text{reh}} = 0.25. \tag{7}$$

The future-feasibility predictor remains frozen and is not differentiated through. At deployment, the aligned generator uses ordinary one-pass diffusion with no branching, predictor query, reranking, or online verifier.

4 RESULTS

4.1 EXPERIMENTAL SETUP

We use the CPCore corpus, constructed from CPSea (Yang et al., 2025), throughout model adaptation and evaluation. After preprocessing, the cyclic-adaptation split contains 70,238 training and 1,498 validation records. Evaluation covers head-to-tail, disulfide, and isopeptide-lactam cyclization under one shared protocol and one frozen joint-feasibility definition.

We compare against PepMDLM and controlled PepTune and TR2-D2 baselines under a shared frozen generator and evaluation protocol; full implementation details are provided in Appendix A.

The evaluation follows the design objective in Section 3.1. We report chemical validity, topology-class realization, joint feasibility, and unique feasible discovery over all terminal outputs. Joint feasibility additionally includes the benchmark closure-geometry criterion. We separately report a stricter topology audit and the post-hoc structural-consistency evaluation in Section 4.4; neither is used during training, search, checkpoint selection, or alignment. Full metric definitions, canonicalization rules, data-split protocol, and compute accounting are provided in Appendix B.

4.2 FUTURE-FEASIBILITY GUIDANCE IMPROVES FEASIBLE DISCOVERY

Table 1(a) shows a substantial improvement in feasible discovery under the shared evaluation protocol. With the same 4,320 terminal evaluations, ReachDiff reaches 77.59% joint feasibility and discovers 3,161 unique jointly feasible products, compared with 59.61% and 2,573 for the strongest common-backbone controller adaptation. Chemical validity and topology-class realization also increase to 81.06% and 78.26%. These gains are consistent with the design of ReachDiff: future-feasibility scores identify promising partial states early, allowing the fixed continuation budget to be concentrated on trajectories more likely to satisfy the coupled constraints.

The improvement is consistent across all three cyclization topologies (Table 1(b)), reaching 83.61% for head-to-tail, 85.07% for disulfide, and 64.10% for isopeptide-lactam generation. This consistency suggests that future-feasibility allocation captures a shared signal of trajectory viability rather than a topology-specific effect.

The advantage persists under matched recorded model work (Table 1(c)). ReachDiff discovers 2,739 unique jointly feasible products, with paired gains of $+231$, $+191$, and $+557$ over PepMDLM, PepTune, and TR2-D2, respectively, with all 95% confidence intervals above zero. Together, these results show that early prediction of terminal success translates into more effective allocation of a fixed inference budget. Additional strict-topology and independent-sampling controls are provided in Appendices D.4 and D.5, respectively.

4.3 WHEN AND WHERE FUTURE-FEASIBILITY INFORMATION APPEARS

4.3.1 TERMINAL DESIGN SUCCESS BECOMES PREDICTABLE EARLY

Terminal success becomes predictable early in generation. Figure 2(a) shows that AUROC rises from 0.798 at 10% diffusion progress to 0.881 at 30%, after which it largely plateaus, creating an early window in which trajectories are already distinguishable while substantial denoising computation remains. Repeated frozen-policy completions confirm that the predicted scores track empirical continuation success (Figure 2(b)). Input ablations further show that this signal is driven primarily by the molecular partial state: a state-only predictor performs nearly as well as the full model, whereas condition and progress alone provide little predictive information.

Table 1: **Main feasible-discovery benchmark.** All methods use the same five-seed evaluation panel. (a) Overall feasibility and unique discovery under 4,320 terminal evaluations; (b) joint-feasible yield by cyclization topology; (c) unique feasible discovery under matched recorded model work. PepTune and TR2-D2 denote common-backbone controller adaptations.

(a) Overall discovery performance

Method	Chemical valid ↑	Topology class realized ↑	Joint feasible ↑	Unique feasible ↑	Unique / model work ↑
PepMDLM	64.70%	57.64%	56.94%	2460	417.17
PepTune	66.69%	60.19%	59.61%	2573	415.15
TR2-D2	64.79%	58.66%	57.99%	2503	362.64
ReachDiff	**81.06%**	**78.26%**	**77.59%**	**3161**	**454.83**

(b) Joint-feasible yield by cyclization topology

Method	H2T ↑	Disulfide ↑	Isopeptide Lactam ↑
PepMDLM	64.86%	60.90%	45.07%
PepTune	65.35%	65.49%	47.99%
TR2-D2	65.35%	62.43%	46.18%
ReachDiff	**83.61%**	**85.07%**	**64.10%**

(c) Matched recorded model work

Method	Unique feasible ↑	ReachDiff − method paired difference (95% CI)
PepMDLM	2508	$+231$ $[9.00, 465.02]$
PepTune	2548	$+191$ $[6.01, 393.00]$
TR2-D2	2182	$+557$ $[383.49, 722.05]$
ReachDiff	**2739**	–

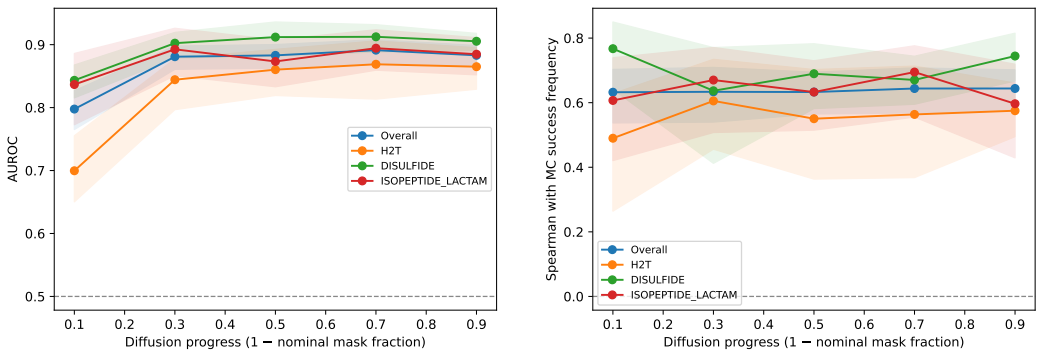


Figure 2: **Future feasibility across diffusion progress.** Left, AUROC for predicting terminal joint feasibility. Right, within-topology Spearman correlation between the future-feasibility score and Monte-Carlo completion success frequency. Shaded regions denote 95% confidence intervals.

The same conclusion holds when scoring time is explicitly synchronized. Under fixed-step scoring, ReachDiff retains a 35.39% gain in unique feasible discovery over matched Random allocation while remaining close to the canonical implementation. Detailed input ablations and synchronized-scoring controls are provided in Appendix C.2.

4.3.2 The gain comes from state-dependent ranking and global competition

Table 2 disentangles the contributions of branching, future-feasibility ranking, and global competition. ReachDiff discovers 31.20% more unique feasible products than Random allocation, while shuffling the predicted scores reduces performance close to random selection, showing that the gain is driven primarily by the learned state-dependent ranking rather than branching alone.

Table 2: **Controlled ablation of future-feasibility-guided selection.** Results use ten fixed search seeds. Random allocation selects survivors randomly under the same branching structure; Shuffled-score allocation permutes future-feasibility scores; Per-parent allocation restricts selection to local sibling competition.

Method	Unique feasible ↑	Feasible yield ↑
PepMDLM	4953	57.36%
Random allocation	4807	57.41%
Shuffled-score allocation	4833	57.75%
Per-parent allocation	5040	58.37%
ReachDiff	**6307**	**77.53%**

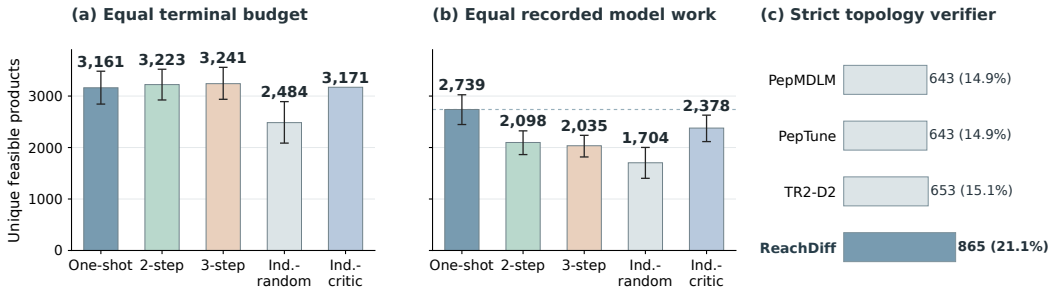


Figure 3: **Allocation-frequency ablation and robustness under a stricter topology verifier. (a)** Repeated guidance provides only modest additional discovery under an equal terminal-evaluation budget. **(b)** Under the common recorded-work budget used in Table 1(c), one-shot allocation provides the strongest compute–discovery trade-off. **(c)** ReachDiff retains its discovery advantage under the stricter topology verifier. Error bars in (a,b) denote 95% percentile bootstrap intervals.

Global competition provides an additional benefit. Restricting the same predictor to per-parent sibling selection reduces unique discovery from 6,307 to 5,040. Per-parent selection reserves capacity for each lineage, whereas global selection can reallocate survival slots toward the highest-future-feasibility trajectories across the entire cohort. This explains why global allocation is particularly effective when continuation quality varies across parents. Complete paired statistics and compute accounting are provided in Appendix D.7.

Selection frequency and candidate construction. Two- and three-step guidance provide only modest gains over one-shot allocation under an equal terminal-evaluation budget, but become less efficient when total recorded model work is matched (Figure 3(a,b)).

Independent-root controls further show that future-feasibility ranking remains effective without shared-prefix branching, while the canonical one-shot procedure retains the best compute efficiency. These results separate the benefit of learned ranking from the efficiency of shared-prefix candidate construction. Detailed accounting and paired bootstrap statistics are reported in Appendix D.2.

4.4 Robustness under stricter topology and structural evaluation

Under a stricter topology definition that additionally enforces exact residue count, terminal/cap annotations, positional anchors, linkage atoms, and the intended ring bond, ReachDiff retains the highest strict-feasible discovery across the same five-seed evaluation panel. As shown in Figure 3(c), ReachDiff reaches 21.1% strict-feasible yield and 865 unique strict-feasible products, compared with at most 15.1% and 653 among the controlled baselines. Additional verifier details are provided in Appendix D.4.

The advantage also transfers to an independent post-hoc structural screen (Table 3). After frozen RDKit/MMFF relaxation, relaxed-feasible yield increases from 42.78% for PepMDLM to 55.56% for ReachDiff, a paired gain of 12.78 percentage points (95% CI $[4.44, 21.39]$). For the standalone

Table 3: **Frozen post-hoc molecular-mechanics consistency screen.** Relaxed-feasible is evaluated after an RDKit/MMFF relaxation procedure excluded from predictor training, search, and alignment.

Path	Method	Relaxed-feasible ↑
Search	PepMDLM	$42.78 \pm 7.66\%$
Search	**ReachDiff**	$\mathbf{55.56 \pm 5.56\%}$
One-pass	Base	$52.16 \pm 4.04\%$
One-pass	TR2-D2 aligned	$56.58 \pm 4.72\%$
One-pass	ReachDiff aligned	$\mathbf{57.36 \pm 5.48\%}$

Table 4: **Secondary transfer from search to a standalone generator.** Panel (a) compares the cost of collecting matched 384-trajectory positive buffers. Panel (b) evaluates the resulting generators using ordinary one-pass diffusion with no search, predictor, reranking, or online verifier. Values are mean ± standard deviation across generation seeds after averaging over three independently trained alignment checkpoints.

(a) Positive-trajectory acquisition

Positive source	Trajectories	Terminal oracles ↓	Backbone rows ↓	Predictor rows	Est. work (10^{15}) ↓
PepMDLM collection	384	762	95,892	0	1.1261
TR2-D2 collection	384	792	112,447	0	1.3581
ReachDiff	384	**570**	**83,951**	1,139	**0.9508**

(b) Search-free one-pass generation

Generator	Chemical@1 ↑	Topology-class@1 ↑	Joint@1 ↑	Diversity ↑
Base	68.83 ± 6.17	61.73 ± 5.43	60.80 ± 5.43	$\mathbf{0.5637 \pm 0.0112}$
PepMDLM aligned	72.43 ± 6.09	67.18 ± 5.26	66.87 ± 5.26	0.5525 ± 0.0068
TR2-D2 aligned	$\mathbf{72.94 \pm 7.59}$	68.21 ± 7.22	67.80 ± 7.41	0.5557 ± 0.0040
ReachDiff aligned	$\mathbf{72.94 \pm 7.92}$	$\mathbf{68.83 \pm 6.08}$	$\mathbf{68.52 \pm 6.17}$	0.5524 ± 0.0024

aligned generator, relaxed-feasible yield increases from 52.16% to 57.36%, corresponding to a paired effect of $+5.195$ pp (95% CI $[0.257, 11.008]$). Together, these results show that the discovery gain persists under both stricter topology criteria and a structural evaluation that is independent of the search objective.

4.5 Search trajectories improve standalone generation

Table 4(a) shows that future-feasibility guidance improves positive-trajectory acquisition efficiency. To collect the same 384 verifier-positive trajectories, ReachDiff requires 570 terminal-oracle calls, compared with 762 for PepMDLM and 792 for TR2-D2, while also using less estimated model work. By concentrating continuation effort on partial states with higher predicted success probability, ReachDiff collects useful training trajectories at lower search cost.

Table 4(b) shows that verifier-positive trajectories from all three collectors provide effective denoising supervision. Under the same alignment schedule, Joint@1 increases from 60.80% for the base generator to 66.87%, 67.80%, and 68.52% for the PepMDLM-, TR2-D2-, and ReachDiff-aligned variants, respectively. Together with the lower acquisition cost in Table 4(a), these results show that ReachDiff provides an efficient route for collecting useful trajectories that can be transferred back into a stronger standalone generator. At deployment, the aligned generator uses ordinary one-pass diffusion without search or predictor evaluation.

4.6 Representative strict-feasible cyclic products

Figure 4 presents representative ReachDiff products across all three cyclization classes. Each passes chemical validity, strict topology, and the independent, frozen MMFF94 structural screen. After relaxation, the corresponding closure bond lengths are 1.383, 2.058, and 1.375 Å for head-to-tail,

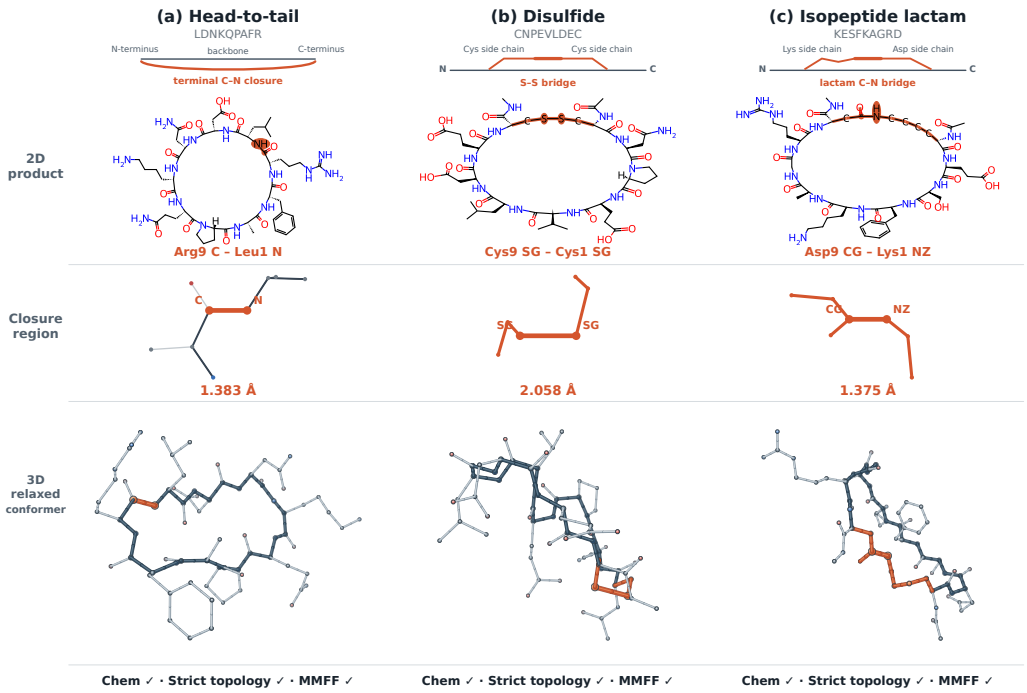


Figure 4: **Representative strict-feasible cyclic products generated by ReachDiff.** Columns show head-to-tail, disulfide, and isopeptide-lactam examples. From top to bottom: 2D molecular structures with the cyclization linkage, relaxed closure regions with bond lengths, and relaxed 3D conformers. All examples pass chemical validity, strict topology, and the frozen MMFF94 structural screen.

disulfide, and isopeptide-lactam cyclization, respectively. These examples illustrate strict-feasible products with well-formed local closure geometry after relaxation.

5 Discussion and Conclusion

The central finding is that cyclic-peptide design success is already partially predictable from unfinished molecular states. This signal emerges early enough in masked diffusion to support a single global allocation of the remaining computation. Matched controls further indicate that the benefit comes from state-dependent ranking and cohort-level competition rather than branching alone. These results suggest that constrained generation can benefit from treating terminal success as a policy-conditional value associated with intermediate states.

Search-discovered trajectories provide a complementary benefit. They can be collected more efficiently and reused as ordinary denoising supervision, allowing the resulting generator to return to standard one-pass diffusion without search or predictor evaluation at deployment. This connects inference-time allocation with a simple route for transferring successful trajectories back into the base generator.

Our conclusions are limited to the evaluated masked-diffusion setting and the frozen operational feasibility criteria used here. The benchmark covers three cyclization classes, while broader generalization across generators, targets, and higher-fidelity physical or experimental evaluation remains to be established. More generally, one-shot allocation is most useful when intermediate states become reliably rankable while substantial continuation computation still remains. Within this regime, our results support a principle: partial-state estimates of terminal success can provide a useful signal for directing finite generation compute toward more promising trajectories.

AI USE STATEMENT

We used generative AI tools to refine the writing and assist with manuscript review. The authors take responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

All headline comparisons use fixed checkpoints, condition panels, canonical deduplication, and pre-specified seeds; the common-backbone benchmark uses five fixed seeds, while the matched allocation-control study uses ten fixed seeds. The appendix specifies predictor construction, search execution, controlled-baseline semantics, topology-class and strict-topology verifier semantics, the independent-sampling replay, geometry audit, positive-trajectory collection, alignment parameters, and statistical pairing; exact frozen configurations and provenance ledgers accompany the submission. Search evaluation is disjoint from predictor training and alignment collection, and the final one-pass evaluation does not invoke search or the predictor.

ETHICS STATEMENT

This work is computational and does not involve human subjects or personal data. Cyclic peptide generation is used as a constrained-generation and hypothesis-generation setting. Computational feasibility does not establish safety, efficacy, or suitability for therapeutic use; any downstream biological application would require dedicated experimental validation and safety review.

REFERENCES

Clément Bechtler and Christina Lamers. Macrocyclization strategies for cyclic peptides and peptidomimetics. *RSC medicinal chemistry*, 12(8):1325–1351, 2021.

Michael Cardei, Jacob K Christopher, Bhavya Kailkhura, Thomas Hartvigsen, and Ferdinando Fioretto. Constrained discrete diffusion. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=Es4s9dtCjR.

Patrick G Dougherty, Ashweta Sahni, and Dehua Pei. Understanding cell penetration of cyclic peptides. *Chemical reviews*, 119(17):10241, 2019.

Xinjian Ji, Alexander L Nielsen, and Christian Heinis. Cyclic peptides for drug development. *Angewandte Chemie International Edition*, 63(3):e202308251, 2024.

Dapeng Jiang, Xiangzhe Kong, Jiaqi Han, Mingyu Li, Rui Jiao, Wenbing Huang, Stefano Ermon, Jianzhu Ma, and Yang Liu. Zero-shot cyclic peptide design via composable geometric constraints. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 27553–27568. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/jiang25d.html.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/d9042abf40782fbce28901c1c9c0e8d8-Abstract-Conference.html.

Xiangzhe Kong, Yinjun Jia, Wenbing Huang, and Yang Liu. Full-atom peptide design with geometric latent diffusion. In *Advances in Neural Information Processing Systems*, volume 37, 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/88ad9774ffcb7a272457e9396f793a07-Abstract-Conference.html.

Gyubin Lee, Truong Nhat Nguyen Bao, Jaesik Yoon, Dongwoo Lee, Minsu Kim, Yoshua Bengio, and Sungjin Ahn. Adaptive inference-time scaling via cyclic diffusion search. *arXiv preprint arXiv:2505.14036*, 2025a.

Jaewoo Lee, Minsu Kim, Sanghyeok Choi, Inhyuck Song, Sujin Yun, Hyeongyu Kang, Woocheol Shin, Taeyoung Yun, Kiyoung Om, and Jinkyoo Park. Diffusion alignment as variational expectation-maximization. In *International Conference on Learning Representations*, volume 2026, pp. 31062–31093, 2026.

Seul Lee, Karsten Kreis, Srimukh Prasad Veccham, Meng Liu, Danny Reidenbach, Yuxing Peng, Saee Paliwal, Weili Nie, and Arash Vahdat. Genmol: A drug discovery generalist with discrete diffusion. *arXiv preprint arXiv:2501.06158*, 2025b.

Jiahan Li, Chaoran Cheng, Zuofan Wu, Ruihan Guo, Shitong Luo, Zhizhou Ren, Jian Peng, and Jianzhu Ma. Full-atom peptide design based on multi-modal flow matching. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 27615–27640. PMLR, 21–27 Jul 2024a. URL https://proceedings.mlr.press/v235/li24o.html.

Xiner Li, Yulai Zhao, Chenyu Wang, Gabriele Scalia, Gokcen Eraslan, Surag Nair, Tommaso Biancalani, Aviv Regev, Sergey Levine, and Masatoshi Uehara. Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding. *arXiv preprint arXiv:2408.08252*, 2024b.

Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, et al. Scaling inference time compute for diffusion models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2523–2534. IEEE, 2025.

Zijing Ou, Chinmay Pani, and Yingzhen Li. Inference-time scaling of discrete diffusion models via importance weighting and optimal proposal design. In *International Conference on Learning Representations*, volume 2026, pp. 36740–36775, 2026.

Jarrid Rector-Brooks, Mohsin Hasan, Zhangzhi Peng, Chenghao Liu, Sarthak Mittal, Nouha Dziri, Michael Bronstein, Pranam Chatterjee, Alexander Tong, and Joey Bose. Steering masked discrete diffusion models via discrete denoising posterior prediction. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/3f2a9f41dd0cb07cd9b60c058c7b1682-Abstract-Conference.html.

Stephen A Rettie, David Juergens, Victor Adebomi, Yensi Flores Bueso, Qinqin Zhao, Alexandria N Leveille, Andi Liu, Asim K Bera, Joana A Wilms, Alina Üffing, et al. Accurate de novo design of high-affinity protein-binding macrocycles using deep learning. *Nature Chemical Biology*, 21(12): 1948–1956, 2025.

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

Yair Schiff, Subham Sahoo, Hao Phung, Guanghan Wang, Sam Boshar, Hugo Dalla-Torre, Bernardo Almeida, Alexander Rush, Thomas Pierrot, and Volodymyr Kuleshov. Simple guidance mechanisms for discrete diffusion models. In *International Conference on Learning Representations*, volume 2025, pp. 43776–43821, 2025.

Maxim Secor, Jovan Damjanovic, Veronika Thost, Sebastian Swanson, Kristine Deibler, and Jesper Ferkinghoff-Borg. Cpgen: Heterochiral cyclic peptide ensemble generation and ensemble-based sequence design. In *The 2026 Workshop on Generative and Agentic AI for Biology*, 2026.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. *arXiv preprint arXiv:2501.06848*, 2025.

Huu Binh Ta, Michael Cardei, Alvaro Velasquez, and Ferdinando Fioretto. Search-augmented masked diffusion models for constrained generation. *arXiv preprint arXiv:2602.02727*, 2026.

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, pp. 59017–59065, 2025a. URL https://arxiv.org/abs/2412.17780.

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

Yitian Wang, Fanmeng Wang, Angxiao Yue, Wentao Guo, Yaning Cui, and Hongteng Xu. Muco: Generative peptide cyclization empowered by multi-stage conformation optimization. *arXiv preprint arXiv:2602.11189*, 2026.

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

Ziyi Yang, Hanyuan Xie, Yinjun Jia, Xiangzhe Kong, Jiqing Zheng, Ziting Zhang, Yang Liu, Lei Liu, and Yanyan Lan. Cpsea: Large-scale cyclic peptide-protein complex dataset for machine learning in cyclic peptide design. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference. Curran Associates, Inc., 2025. doi: 10.52202/085713-1479. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/3eec719ab86712d32b065c5977f94ad0-Paper-Datasets_and_Benchmarks_Track.pdf.

Jingjie Zhang, Hanqun Cao, Haosen Shi, He Mutian, Yu Wang, Zijun Gao, Fang Wu, Xiaojun Yao, Chang-Yu Hsieh, Sinno Jialin Pan, et al. Geocycler: Reward-aligned 3d diffusion for constraint-conditioned cyclic peptide design. *arXiv preprint arXiv:2605.23407*, 2026a.

Xiangcheng Zhang, Haowei Lin, Haotian Ye, James Y Zou, Jianzhu Ma, Yitao Liang, and Yilun Du. Inference-time scaling of diffusion models through classical search. In *International Conference on Learning Representations*, volume 2026, pp. 39133–39166, 2026b.

Yifan Zhao, Lang Qin, and Jintai Chen. Apcyc: Property-informed design of cyclic peptides via automated cyclization. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 13020–13031, 2026.

Xiangxin Zhou, Mingyu Li, Yi Xiao, Jiahan Li, Dongyu Xue, Zaixiang Zheng, Jianzhu Ma, and Quanquan Gu. Designing cyclic peptides via harmonic SDE with atom-bond modeling. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 78827–78858. PMLR, 13–19 Jul 2025. URL https://proceedings.mlr.press/v267/zhou25i.html.

Alessandro Zorzi, Kaycie Deyle, and Christian Heinis. Cyclic peptide therapeutics: past, present and future. *Current opinion in chemical biology*, 38:24–29, 2017.

APPENDIX

A CONTROLLED BASELINE IMPLEMENTATIONS

PepMDLM / independent sampling. PepMDLM uses the same frozen topology-conditioned cyclic generator as as ReachDiff and performs ordinary independent terminal sampling without branching, intermediate selection, predictor scoring, or reranking. All terminal attempts, including invalid and duplicate outputs, count toward the evaluation budget; products passing the frozen verifier are retained and canonically deduplicated. This arm therefore measures the base generation performance and serves as the independent-sampling/rejection-sampling control under the shared protocol.

Controlled search-controller adaptations. PepTune and TR2-D2 use the same frozen generator and binary joint-feasibility verifier as ReachDiff, but differ in their search controllers. For both adaptations, terminal rollouts used during search are included in the shared terminal-verifier budget and feasibility denominator; no separate post-search output set is used. These baselines adapt the corresponding search principles to the same cyclic-feasibility setting for controlled comparison. All controllers use the same frozen generator, generation conditions, and terminal evaluation protocol.

ReachDiff. ReachDiff performs a single branching stage followed by global top-B selection using the frozen future-feasibility predictor. Selected trajectories then resume ordinary sampling under the unchanged base generator. Unlike the controlled tree-search baselines, the predictor predicts future terminal feasibility from partial states and does not invoke the terminal verifier before completion.

B EVALUATION METRICS

Chemical validity. A terminal output is chemically valid if the reconstructed molecular graph passes the fixed parsing, sanitization, connectivity, valence, and peptide-backbone checks used throughout evaluation.

Topology-class realization. Topology-class realization measures whether the generated molecular graph realizes the requested head-to-tail, disulfide, or isopeptide-lactam closure class.

Joint feasibility. Joint feasibility follows Eq. 1 and requires chemical validity, realization of the requested topology class, and the benchmark closure-geometry criterion.

Strict-topology audit. The strict-topology audit applies a more exact structural definition of the requested cyclization, including residue count, termini and caps, positional anchors and linkage atoms, and formation of the intended ring bond. This audit is performed only after generation and is not used for predictor training, search, checkpoint selection, or generator alignment.

Structural-consistency audit. Section 4.4 evaluates generated products using an independently frozen post-hoc relaxation protocol. This evaluation is likewise excluded from the search objective and from all model-selection decisions.

Unique feasible discovery. Feasible terminal products are deduplicated by canonical isomeric molecular identity with stereochemistry retained. Thus, unique feasible discovery counts distinct molecular products rather than repeated generations of the same product.

Compute accounting. Terminal evaluations and recorded model work are reported separately. Recorded model work additionally accounts for unfinished generator execution and predictor evaluation, and is used only for matched-work comparisons.

Data separation. Predictor states are split at the level of complete terminal rollouts, so partial states originating from the same rollout never cross training, validation, and test partitions. The primary search evaluation is disjoint from predictor training.

Model-work efficiency. Recorded model work is computed from the fixed execution ledger and accounts for paid generator and predictor execution. It is used for relative execution accounting and is reported separately from hardware FLOPs or wall-clock time. We report both unique feasible products per model work and a fixed-work comparison in which all methods are truncated under the same recorded-work budget.

C Supplementary Experimental Details

C.1 Predictor data and architecture

The predictor dataset is collected from ordinary sampling by the frozen step-10k cyclic generator. It contains 6,000 terminal trajectories, 2,000 per topology. Partial states are retained near the five pre-specified mask levels $0.9, 0.7, 0.5, 0.3, 0.1$, producing 30,000 observations. Each partial state inherits the binary terminal joint-feasibility outcome of its source trajectory. The split unit is the terminal rollout: all partial states derived from the same rollout are assigned together to train, validation, or test. The audited split contains 18,000/6,000/6,000 partial-state observations corresponding to 3,600/1,200/1,200 complete terminal trajectories, with no source rollout crossing split boundaries. Predictor training uses seed 11731 and shuffle seed 11732. A separate Monte-Carlo analysis evaluates 1,500 partial states with 16 frozen-policy completions each, for 24,000 completions total.

The predictor feature extractor mean-pools the frozen RoFormer final-layer hidden states over content positions, yielding a 768-dimensional representation. A 16-dimensional topology embedding and a 16-dimensional embedding of realized mask fraction and diffusion grid time are concatenated and passed through an approximately $800 \to 128 \to 1$ MLP. The default predictor contains approximately 110,945 parameters and is trained with BCE-with-logits. The reported predictor uses this single scalar head throughout all main experiments.

C.2 State-information and synchronized-scoring controls

The full predictor observes both a pooled molecular-state representation $h(x_t)$ and low-dimensional variables (c, m_t, t). We compare three predictors trained on the same audited terminal-rollout split and labels: (i) condition/progress only, $q(c, m_t, t)$; (ii) molecular state only, $q(h(x_t))$; and (iii) the full predictor, $q(h(x_t), c, m_t, t)$. Within-cohort ranking compares candidates from the same search cohort, while the Monte-Carlo gap reports the difference in empirical completion frequency between the highest- and lowest-scoring 20% of states.

Table 5: **Predictor input ablation.** All variants use the same terminal-rollout split and bootstrap 95% confidence intervals. Within-cohort ranking is evaluated on 27 cohorts with 12 candidates per cohort and 16 frozen-policy completions per candidate.

Predictor input	AUROC ↑	MC Spearman ↑	Within-cohort rank acc. ↑	Top–bottom MC gap (pp) ↑
Condition/progress only	0.5373 [0.4916, 0.5793]	0.0982 [0.0094, 0.1788]	0.5015 [0.4494, 0.5623]	9.13 [−3.03, 20.33]
State only	0.8683 [0.8526, 0.8829]	0.6316 [0.5933, 0.6680]	0.7734 [0.7177, 0.8295]	85.00 [78.70, 88.69]
Full predictor	**0.8702** [0.8542, 0.8843]	**0.6374** [0.5975, 0.6728]	**0.7741** [0.7190, 0.8265]	84.25 [78.64, 88.64]

Relative to condition/progress-only input, the full predictor improves AUROC by $+0.3329$ (95% CI $[0.2888, 0.3791]$) and within-cohort ranking accuracy by $+27.26$ percentage points ($[19.68, 34.88]$). The state-only predictor already recovers nearly the full performance, showing that the frozen partial-state representation contains the dominant future-feasibility signal.

We separately synchronize the search scoring time to remove first-crossing timing variation. All methods below use the same five seeds (2026091701–2026091705), 4,320 terminal outputs, branching opportunities, candidate counts, and terminal budget. Fixed-step candidates are scored at nominal step 377 ($t \approx 0.3984$).

ReachDiff with threshold scoring exceeds canonical Random allocation by $+775$ unique feasible products ($+32.48\%$, paired 95% CI $[616.00, 923.00]$). Under synchronized scoring, ReachDiff exceeds fixed-step Random allocation by $+837$ products ($+35.39\%$, $[672.99, 990.00]$). Fixed-step and canonical ReachDiff differ by only $+41$ products ($+1.30\%$, $[-17.68, 98.00]$), preserving the same discovery level when scoring time is held constant.

Table 6: **Fixed-step search control.** Synchronized scoring preserves the future-feasibility-guided discovery gain under a common nominal diffusion time.

Method	Unique feasible (95% CI) ↑	Feasible yield (95% CI) ↑
PepMDLM	2460 [2029.98, 2906.00]	56.94% [46.99, 67.27]
Threshold-scored random allocation	2386 [1980.00, 2799.00]	57.15% [46.80, 67.66]
ReachDiff with threshold scoring	3161 [2844.16, 3482.67]	77.59% [68.26, 87.22]
Synchronized random allocation	2365 [1947.98, 2805.03]	56.53% [46.06, 67.64]
ReachDiff with synchronized scoring	**3202** [2885.50, 3531.01]	**78.47%** [69.21, 88.38]

C.3 Exact search execution

The reported search uses six all-MASK roots. Each root advances to the first realized state with mask fraction at most 0.70. If the realized mask fraction remains above 0.65, the parent is forked into two continuations; otherwise one direct candidate is retained. All candidates then advance to the first realized state with mask fraction at most 0.65, after which the frozen future-feasibility scores the entire cohort and the global top six survive. Because first-crossing states can differ slightly in realized mask fraction, m_t and t are included explicitly in the predictor input. Appendix C.2 shows that condition/progress-only ranking is near chance and that synchronized fixed-step scoring preserves the discovery gain. Selected trajectories then complete ordinary diffusion and final argmax. Each cohort therefore ends with six terminal verifier evaluations. The exact number of generator calls depends on realized mask crossings, cache misses, padding, and batch composition, so compute comparisons use measured execution ledgers.

C.4 Pseudocode for cohort allocation

Algorithm 1 Future-feasibility-guided cohort allocation

Require: Frozen generator π_{θ_0}, frozen predictor q_ϕ, condition c, survival budget B
1: Initialize B independent all-MASK sampler replicas
2: Advance all replicas to the warmup threshold m_{warm}
3: Apply the fixed branching rule to construct candidate cohort $\mathcal{C}$
4: Advance all candidates to the scoring threshold m_{score}
5: Compute $q_\phi(x, c, m_t, t)$ for every $x \in \mathcal{C}$
6: Retain the global top-B candidates
7: Complete retained states using ordinary sampling from π_{θ_0}
8: Apply the terminal verifier and canonically deduplicate feasible products

C.5 Controlled baseline execution

PepTune and TR2-D2 search are evaluated through the project-frozen controlled implementations in the common topology-conditioned masked-diffusion harness. All four arms share the same condition panels, terminal joint-feasibility verifier, canonical product identity, and equal-terminal or fixed recorded-work budget used by the corresponding comparison. These controlled implementations are distinguished in the text from end-to-end reproduction of the original task objectives. The frozen configurations and execution ledgers accompanying the submission record the exact branching, selection, and stopping settings used for each baseline.

C.6 Verifier semantics and canonical identity

Joint feasibility is Eq. 1. Because this verifier defines both predictor supervision and the primary endpoint, its implementation is benchmark-defining and is specified explicitly below. Chemical validity uses RDKit parsing, sanitization, valence checks, connectivity, and peptide-backbone rules. Topology success asks whether graph reconstruction yields the requested head-to-tail, disulfide, or isopeptide-lactam class. The geometry term uses the fixed benchmark surrogate with conformer generation, closure-distance, severe-clash, and force-field single-point-energy criteria. Anchor, cap, and residue-count fields are recorded as supplementary diagnostics alongside the main topology label.

All verifier thresholds, conformer-generation settings, clash criteria, and force-field cutoffs are frozen across methods and are provided in the anonymized evaluation configuration accompanying the submission.

For discovery counting, atom-map annotations are removed and products are deduplicated by RDKit canonical isomeric SMILES with stereochemistry retained. Thus, "unique" is defined at the level of canonical molecular products with retained stereochemistry.

C.7 Positive-trajectory alignment

The ReachDiff positive pool is collected on conditions separate from primary one-pass evaluation. First-positive family quotas retain 128 successful trajectories per topology, for 384 trajectories total; these correspond to 365 unique terminal products. Each update samples real intermediate states from positive trajectories and mixes them with rehearsal examples from the cyclic training split. The trainable subset contains approximately 15.56M parameters, including Q/V LoRA components across eight encoder layers, existing fully trainable parameters in the final two layers, topology embeddings, a condition adapter, and topology-vocabulary bias terms. Remaining pretrained parameters stay frozen.

The PepMDLM-positive and TR2-D2 alignment controls use the same pool size, topology quotas, training seeds, 500 updates, optimizer, batch composition, rehearsal coefficient, and trainable parameter subset. The variants differ only in the source of successful trajectories, and collection compute is measured directly from the execution ledger under each collector. PepMDLM collection yields 384 unique terminal products from 762 oracle calls and 95,892 backbone rows; TR2-D2 collection yields 384 unique products from 792 calls and 112,447 backbone rows; ReachDiff yields 365 unique products from 570 calls, 83,951 backbone rows, and 1,139 predictor rows. The corresponding estimated work values are 1.1261, 1.3581, and 0.9508×10^{15}, respectively.

Alignment uses training seeds 8101, 8102, and 8103, with 500 AdamW updates at learning rate 10^{-5}, batch composition 6 positive + 6 rehearsal examples, $\lambda_{\mathrm{reh}} = 0.25$, and 15,557,601 trainable parameters. One-pass evaluation uses generation seeds 2041103501–2041103503. Base contributes 324 outputs, while each aligned family contributes 972 outputs across the three trained checkpoints.

C.8 Statistical pairing and aggregation

All paired search comparisons use the same fixed conditions and generation seeds within a method pair. Confidence intervals are computed on paired per-seed discovery differences, making the generation seed the statistical pairing unit. For predictor evaluation, all partial states originating from the same terminal rollout must remain in the same dataset split, as noted above.

For one-pass aligned-model evaluation, the experimental hierarchy contains three independently trained alignment checkpoints and three generation seeds. For each generation seed, metrics are first averaged across the three training checkpoints; the reported mean and standard deviation are then computed across the three generation seeds. Paired confidence intervals in the alignment comparison are computed by the audited evaluation pipeline on the corresponding paired aggregated metrics.

D Additional audited results

D.1 Full temporal future-feasibility analysis

Table 7 reports all five pre-specified mask-level bins used by the temporal-evaluation pipeline. We report terminal-label AUROC and Monte-Carlo success-frequency correlation separately: the former evaluates binary terminal outcomes, whereas the latter evaluates ranking against repeated continuation success frequencies. The early AUROC increase from mask fraction 0.9 to 0.7 is $+0.083$ $[0.055, 0.114]$; the later change from 0.7 to 0.1 is $+0.002$ $[-0.010, 0.016]$.

D.2 Allocation-frequency and candidate-construction audit

Figure 3(a,b) separates the raw benefit of repeated selection from its additional inference cost. Under the equal-terminal setting, all allocation-frequency variants are compared using the same number

Table 7: Temporal future-feasibility diagnostics across the five pre-specified progress bins.

Progress	AUROC (95% CI)	MC Spearman	Top20–bottom20 MC success
0.1	0.798 [0.765,0.827]	0.632	+73.02 pp
0.3	0.881 [0.861,0.898]	0.634	+83.13 pp
0.5	0.883 [0.863,0.900]	0.633	+89.48 pp
0.7	0.891 [0.871,0.908]	0.644	+83.23 pp
0.9	0.883 [0.867,0.899]	0.644	+85.00 pp

Table 8: **Allocation-frequency and candidate-construction audit under matched recorded model work.** All variants are chronologically truncated at the same recorded-work cap used in Table 1(c), $5.963821554773376 \times 10^{15}$ recorded operations. Intervals are 95% percentile bootstrap intervals obtained from 5,000 paired resamples over the five fixed generation seeds and condition blocks. The final column reports paired differences relative to the canonical one-shot procedure.

Method	Unique feasible ↑	Terminal outputs	Difference vs. one-shot
One-shot	2739 [2447, 3025]	3720	–
2-step	2098 [1863, 2324]	2784	-641 [-744, -508]
3-step	2035 [1817, 2237]	2664	-704 [-807, -590]
Ind.-random	1704 [1402, 2002]	2934	-1035 [-1154, -910]
Ind.-critic	2378 [2115, 2629]	2940	-361 [-480, -255]

of completed terminal evaluations. Two- and three-step guidance yield only modestly more unique feasible products than the canonical one-shot procedure, indicating that repeated intervention can provide a small selection benefit when its additional intermediate computation is not charged against the terminal budget.

To evaluate the corresponding compute–discovery trade-off, we additionally truncate all variants at the same recorded-model-work cap used in Table 1(c), $5.963821554773376 \times 10^{15}$ recorded operations. Chronological truncation is applied directly to the saved execution ledgers; no molecules are regenerated and no search outputs are reordered. Because repeated-guidance variants spend additional computation on intermediate generation and predictor evaluation, equal recorded work does not imply an equal number of completed terminal trajectories.

Under this common work budget, one-shot allocation discovers 2,739 unique feasible products, compared with 2,098 and 2,035 for the two- and three-step variants, respectively. The corresponding paired differences are -641 and -704, with both 95% bootstrap intervals excluding zero. Repeated intervention therefore provides a small raw-discovery gain under an equal-terminal budget, but that gain does not compensate for the additional intermediate generation and scoring required when total recorded work is fixed. The one-shot procedure consequently provides the more favorable compute–discovery trade-off.

The independent-root controls further separate future-feasibility ranking from candidate construction. Under matched recorded work, independent random and critic-based selection complete nearly identical numbers of terminal trajectories (2,934 and 2,940), while future-feasibility ranking increases unique feasible discovery from 1,704 to 2,378. Thus, the ranking signal remains effective even when the shared-prefix branching structure is removed. However, the canonical one-shot procedure reaches 2,739 unique feasible products and completes 3,720 terminal trajectories under the same total-work budget. This comparison indicates that shared-prefix branching is not required for the ranking signal itself, but provides a more compute-efficient candidate construction for exploiting that signal.

D.3 Verifier semantics audit

The historical topology label denotes cyclization-class realization rather than a full positional-topology annotation; stricter topology diagnostics are therefore evaluated separately. The frozen geometry surrogate remains part of the benchmark verifier and is applied with fixed thresholds across all methods.

D.4 STRICT-TOPOLOGY ROBUSTNESS

We additionally evaluate all saved terminal outputs under a stricter topology definition that enforces exact residue count, terminal/cap annotations, positional anchors and linkage atoms, and formation of the intended ring bond. This replay uses the same five equal-terminal-budget seeds as the main benchmark, with 4,320 terminal outputs per method, and does not regenerate molecules or modify any verifier threshold.

Table 9: **Five-seed robustness under the stricter topology verifier.** Strict feasibility requires chemical validity and the stricter topology definition in addition to the frozen benchmark verifier.

Method	Strict feasible ↑	Unique strict-feasible ↑
PepMDLM	14.9%	643
PepTune	14.9%	643
TR2-D2	15.1%	653
ReachDiff	**21.1%**	**865**

Across the five-seed evaluation panel, ReachDiff reaches 21.1% strict-feasible yield, corresponding to 910 of 4,320 terminal outputs, and produces 865 unique strict-feasible products. The strongest controlled baseline reaches 15.1% strict feasibility with 653 unique products. Thus, the discovery advantage observed under the main benchmark remains present when the output criterion is strengthened to require the stricter positional-topology constraints.

D.5 INDEPENDENT-SAMPLING AND REJECTION-SAMPLING CONTROL

The PepMDLM arm is ordinary independent sampling from the same frozen generator: six independent masked roots per cohort, no predictor calls, no branching/search/survivor selection, and no model-based reranking. The rejection rule simply retains terminal products passing the frozen verifier; every attempt, including invalid and duplicate outputs, is charged to the budget. Table 10 gives the first-seed accounting.

Table 10: **Independent-sampling control on the first seed.** Equal-oracle uses 864 terminal attempts per method; matched work truncates each chronological execution ledger at approximately the same recorded-work cap.

Metric	Independent sampling	ReachDiff
Equal-oracle terminal attempts	864	864
Equal-oracle joint feasible	55.0%	**76.5%**
Equal-oracle strict feasible	15.6%	**20.3%**
Equal-oracle unique feasible / strict feasible	475 / 135	**625 / 165**
Matched-work terminal attempts	**882**	756
Matched-work estimated work (10^{15})	1.196	1.213
Matched-work unique feasible / strict feasible	486 / 140	**545 / 143**
Unique feasible per 10^{15} recorded work	406.2	**449.3**

The first-seed strict-feasible matched-work difference is deliberately not treated as a significance result. More importantly, under the five-seed matched-work chronological replay, the two methods have nearly identical recorded work (relative gap 0.00038%). After truncation at the common recorded-work cap, independent sampling yields 2,508 unique topology-class-feasible and 654 unique strict-feasible products, whereas ReachDiff yields 2,739 and 755, respectively. These matched-work counts are distinct from the equal-terminal strict-topology results reported in Table 9.

D.6 Representative strict-feasible case-study details

The three main-text cases come from generation seed 2066091701 and were evaluated with the same frozen MMFF94 three-conformer post-hoc pipeline. Table 11 records the molecule-level diagnostics used in Figure 4.

Table 11: **Representative strict-feasible cyclic products.** Residue numbering is one-based. All three cases pass chemical validity, topology-class realization, strict topology, and the frozen MMFF94 structural screen, with zero severe nonlocal heavy-atom clashes.

Case	Sequence	Cyclization linkage	Relaxed closure	Severe clashes	MMFF energy / heavy atom
H2T	LDNKQPAFR	Arg9 C–Leu1 N	1.383 Å	0	-2.022
Disulfide	CNPEVLDEC	Cys9 SG–Cys1 SG	2.058 Å	0	-0.792
Isopeptide/lactam	KESFKAGRD	Asp9 CG–Lys1 NZ	1.375 Å	0	-1.828

The H2T and disulfide cases reuse completed physical-screen verdicts. The lactam example was selected from saved strict-positive products using a predeclared pre-relaxation diagnostic ranking and then passed through the same frozen pipeline. These examples are used only for molecule-level visualization and do not define a population-level physical success rate.

D.7 Compute and efficiency audit

The main common-backbone comparator study uses the same five fixed seeds in both the equal-terminal-evaluation and matched recorded-work analyses. Under the fixed recorded model-work cap, the unique-feasible counts are 2,508 for PepMDLM, 2,548 for PepTune, 2,182 for TR2-D2, and 2,739 for ReachDiff. The corrected paired differences are

$$\begin{aligned} ReachDiff - \text{PepMDLM}: &\quad +231\ [9.00, 465.02], \\ ReachDiff - \text{PepTune}: &\quad +191\ [6.01, 393.00], \\ ReachDiff - \text{TR2-D2}: &\quad +557\ [383.49, 722.05]. \end{aligned}$$

Thus, the matched-work advantage is statistically resolved against all three common-backbone comparators.

The ten-seed matched-control study separately tests search allocation under matched branching. Backbone-work-normalized unique discovery for ReachDiff is 29.84% higher than Random allocation with 95% CI $[21.07, 41.08]\%$. The accounting includes paid unfinished backbone work and predictor execution. It is an execution-ledger metric; direct hardware-FLOP and wall-clock measurements are left to systems-level profiling.

D.8 Search diversity and novelty

Search-time Morgan diversity is 0.5470 ± 0.0021 for ReachDiff and 0.5549 ± 0.0022 for PepMDLM. The paired difference is $-0.008\ [-0.009, -0.007]$. This corresponds to a small but measurable diversity reduction alongside the larger gain in feasible discovery. The audited novelty check finds no exact canonical-product replication against the comparison corpus; similarity-based analyses further characterize motif-level overlap.

D.9 Frozen post-hoc molecular-mechanics consistency protocol and outcomes

The held-out structural-consistency evaluation uses a frozen RDKit/MMFF-based post-hoc relaxation workflow stored with the supplementary configuration and protocol artifact. The protocol is fixed before the expanded evaluation, and no generated output is removed because relaxation or conformer construction fails. Reported failure accounting includes parsing, stereochemistry, convergence, conformer, topology/closure, and other failure categories. The primary endpoint is all-output relaxed-feasible yield.

The expanded search evaluation contains 360 selected outputs per method. The one-pass evaluation contains 648 Base outputs and 1,944 outputs per aligned method, using six pre-fixed generation seeds and all three fixed training checkpoints. Search relaxed-feasible rates are $42.78 \pm 7.66\%$

for PepMDLM and $55.56 \pm 5.56\%$ for ReachDiff. One-pass rates are $52.16 \pm 4.04\%$ for Base, $56.58 \pm 4.72\%$ for TR2-D2 aligned, and $57.36 \pm 5.48\%$ for ReachDiff aligned. The overall one-pass ReachDiff–Base effect is $+5.195$ pp with a 95% CI of $[0.257, 11.008]$ under the frozen post-hoc screen.

D.10 SECOND-CHECKPOINT REPLICATION

The additional 5k cyclic checkpoint was selected before observing new outcomes. Its scalar predictor uses the same architecture and training definition as the 10k study while the generator backbone remains frozen. The 5k predictor reaches AUROC 0.874 $[0.857, 0.891]$. Within-checkpoint relative unique-feasible gains are 36.41% (95% CI $[25.83, 51.69]\%$) over Random allocation and 35.97% ($[26.57, 48.67]\%$) over shuffled-score selection. The gain over per-parent selection is 28.50% ($[19.27, 42.31]\%$). This replication provides a second within-family checkpoint confirmation of the future-feasibility effect.

D.11 NEGATIVE-AWARE DISTILLATION

The default generator alignment is positive-trajectory denoising plus rehearsal. Negative outcomes already supervise the future-feasibility predictor. The canonical alignment uses positive-trajectory denoising plus rehearsal. Negative-aware and preference-based generator alignment are natural extensions for future study.