CPGEN: Heterochiral Cyclic Peptide Ensemble Generation and Ensemble-Based Sequence Design

Anonymous Authors[1]

Abstract

We present CPGEN, which adapts the *La-Proteina* partially latent flow matching framework for heterochiral cyclic peptide conformer ensemble generation and ensemble-based sequence design. A variational autoencoder compresses all-atom structures into a per-residue latent space, and a flow matching denoiser generates in the joint C_α + latent variable space, conditioned on amino acid sequence, C_α coordinates, or per-residue ϕ,ψ-backbone distribution via multi-modal classifier-free guidance. CPGEN is trained on $\sim$7M Rosetta-sampled conformers spanning $\sim$7,000 heterochiral sequences containing 39 amino acid types (19 L-amino + 19 D-amino + Glycine). The model achieves: (1) sequence-conditioned ensemble generation with ϕ,ψ-backbone distribution similarity to the target distribution; (2) ensemble-based sequence design where conditioning on ϕ,ψ-backbone distributions recovers $\sim$90% of residue identities (100% for 12-mers), establishing that conformational ensemble distributions can be used for sequence inference; and (3) native D-amino acid support, with correct mirror-image backbone preferences emerging from model training without explicit chirality-aware design choices.

1. Introduction

Cyclic peptides occupy a unique niche in therapeutic design, combining the target specificity of biologics with the cell permeability of small molecules while also providing favorable resistance to proteolysis (Xiao et al., 2025; Wang et al., 2022; Muttenthaler et al., 2021; Bhardwaj et al., 2016; White et al., 2011). However, realizing their therapeutic potential poses several computational challenges. First, cyclic peptides have multiple thermally-accessible conformations and the relative populations of these conformational states, not any single structure, govern binding affinity, selectivity, membrane permeability and many other biophysical properties important for therapeutic developability (Ge et al., 2021; Calvo-Barreiro et al., 2025). This is particularly evident in the property of *chameleonicity* where a cyclic peptide dynamically adopts different conformations with shielded or exposed hydrogen bond donors or acceptors to modulate the membrane permeability depending on its environment (Ono et al., 2021; Danelius et al., 2020; Linker et al., 2023; Poongavanam et al., 2024). This changes the paradigm of the structure-activity relationship to an ensemble-activity relationship. Second, therapeutically relevant cyclic peptides frequently include nonstandard amino acids, such as D- and N-methylated (NMe) amino acid, to improve metabolic stability and pharmacokinetics. While therapeutically beneficial, these modifications dramatically expand the molecular design space, posing additional challenges for both structure prediction and sequence design (Hosono et al., 2023; A. Viarengo-Baker et al., 2021).

Ensemble generation is expensive. Currently, molecular dynamics with enhanced sampling is the most effective and robust manner of ensemble prediction for cyclic peptides but is also prohibitively expensive where microsecond-scale simulations are typically required to cover the rugged energy landscape of constrained peptides (Damjanovic et al., 2021; Laio & Parrinello, 2002; Koch et al., 2024). Rosetta-based conformer sampling (Leman et al., 2020; Bhardwaj et al., 2016) is faster but still requires seconds per conformer at reduced accuracy, making the generation of $\sim$1,000-member ensembles (needed for statistical analysis) feasible in principle, but impractical for high-throughput screening campaigns. Recent machine learning methods have begun to address this bottleneck, demonstrating promising results for conformational ensemble prediction and generation of these ensembles with limited scope. *The need for a fast and accurate all-atom cyclic peptide ensemble generator still exists*.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Submitted to the Generative and Agentic AI for Biology Workshop @ ICML 2026. Do not distribute.

Ensemble-based sequence design is lacking. Inverse folding methods such as ProteinMPNN (Dauparas et al., 2022) and ESM-IF (Hsu et al., 2022) predict sequences from single backbone structures, while methods for designing sequences from a conformational ensemble are lacking. This is a significant challenge for cyclic peptides as the ensemble is a crucial biologically relevant object during design. A method that maps ensemble distributions to sequences would enable a new design paradigm. Namely, *the specification of the desired conformational behavior and then the prediction of what sequence produces it*.

D-amino acid support is limited. Despite their routine incorporation in therapeutic cyclic peptides for improved proteolytic stability and to access positive-ϕ Ramachandran regions essential for ring closure (Bhardwaj et al., 2016), D-amino acids remain underrepresented in structure generation and sequence design methods. While some predictive and generative models (Feller & Wilke, 2024; Feller et al., 2026; Li et al., 2025; Abramson et al., 2024; Zhang et al., 2025; Mao et al., 2025) have begun to include D-amino acids, they largely remain underappreciated and are often incorporated into models post hoc. We believe that *the therapeutic relevance of* D*-amino acids requires them to be placed on an equal footing with their* L*-amino acid counterparts*.

We address these challenges with CPGEN, adapting the partially latent flow matching framework of *La-Proteina* (Geffner et al., 2025) for cyclic peptide ensemble generation:

1. **Fast ensemble generation from sequence.** Flow matching in the joint C_α + latent space generates 1,000-conformer ensembles in 12-25 seconds on a single GPU (80–1,200$\times$ faster than Rosetta), achieving 0.52 Ramachandran cosine similarity, enabling high-throughput screening at previously inaccessible scales.
2. **Ensemble-based sequence design.** Multi-modal classifier-free guidance enables conditional generation on a variety of signals. Conditioning on ϕ,ψ-backbone distributions recovers 90% of individual residue identities (100% for 12-mers), demonstrating sequence inference capabilities from ensemble descriptions instead of single, static backbone structures.
3. **Native D-amino acid support.** A 39-type vocabulary (20 L + 19 D) with variable-length architecture (6-12 residues) handles heterochiral peptides on equal footing with homochiral peptides, with demonstrated correct mirror-image backbone distribution generation.

2. Related Work

Protein structure and ensemble generation. Flow matching (Lipman et al., 2022) and diffusion models have been designed to generate proteins and protein backbones (Watson et al., 2023; Ingraham et al., 2023; **?**; Bose et al., 2024; Abramson et al., 2024; Passaro et al., 2025), while recently the partially latent framework of *La-Proteina* (Geffner et al., 2025) jointly generates sequence and all-atom structure. Models for conformer ensemble generation include Boltzmann generators (Noé et al., 2019) which learn invertible maps to equilibrium distributions and AlphaFlow (Jing et al., 2024) which fine-tuned AlphaFold2 with a flow matching objective in pairwise distance space to sample protein conformational ensembles. Other ensemble generators include ESMAdam (Yu et al., 2025) and BioEmu (Lewis et al., 2024). These models replace expensive molecular dynamics using various approaches, but all assume linear protein chain topology and none addresses multi-modal conditioning or heterochirality.

Protein inverse folding. Inverse folding methods such as ProteinMPNN (Dauparas et al., 2022), ESM-IF (Hsu et al., 2022), LigandMPNN (Dauparas et al., 2025), PiFold (Gao et al., 2023) predict sequences from single backbone structures, while DynamicMPNN(Abrudan et al., 2025) predicts sequences from multiple structures. Some work has been done in equilibrium property sequence design for intrinsically disordered proteins (Krueger et al., 2025).

Cyclic peptide ensemble prediction and generation. Miao & Lin introduced StrEAMM (Miao et al., 2021) for heterochiral cyclic peptide ensemble prediction, where models made static-length predictions of backbone conformer populations across conformer bins; Hui et al. generalized this to more non-canonical amino acids using physics-based featurization (Hui et al., 2025). RINGER (Grambow et al., 2024) applied a BERT-based DDPM operating in internal coordinate space (bond distances, angles, dihedrals) to generate conformer ensembles for macrocyclic peptides, though it requires a separate constrained optimization for ring closure reconstruction. Bayaraa et al. (Bayaraa et al., 2026) demonstrated that diffusion models can produce simulation-quality conformer ensembles for heterochiral cyclic peptides, but without side-chain generation, conditional design, or variable-length support.

CPGEN addresses these gaps by combining partially latent flow matching with multi-modal classifier-free guidance (Ho & Salimans, 2022) over sequence, structure, and ϕ,ψ-backbone distributions conditioning signals.

3. Methods and Dataset

3.1. Problem Setting

Given a cyclic peptide with n residues ($6 \leq n \leq 12$, potentially including D-amino acids), we generate an ensemble of all-atom conformers by repeated sampling in the joint space

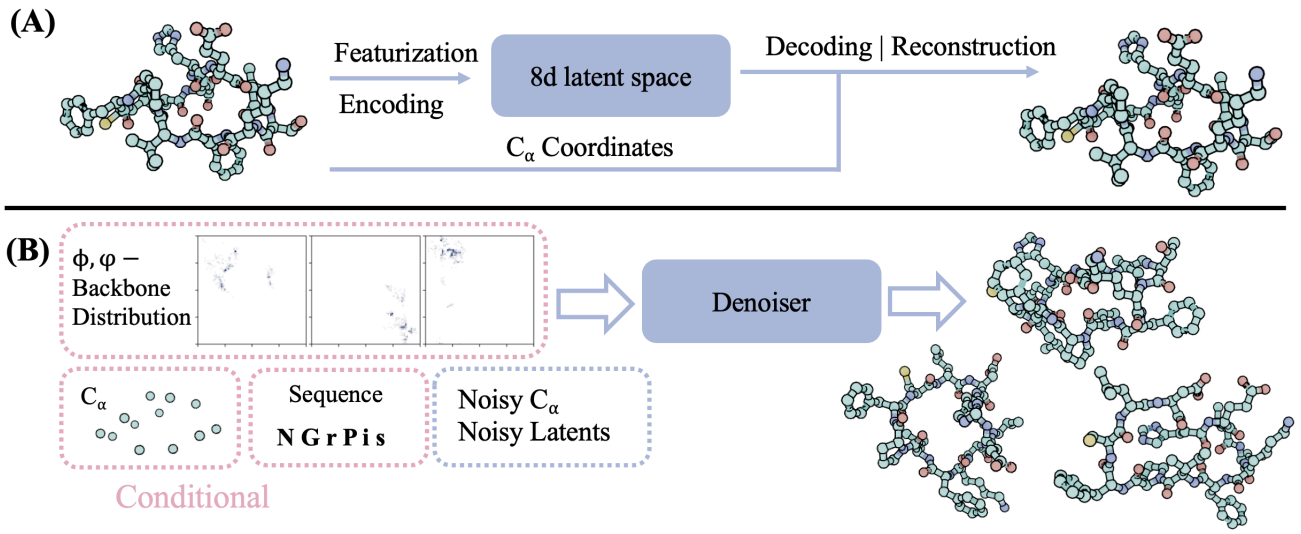

Figure 1. CPGEN consists of an encoder and decoder (A), and joint denoiser (B). The encoder predicts per-residue latent variables while the decoder outputs sequence and non-C_α coordinates from latent variables and C_α coordinates allowing for atomistic reconstruction. To generate sequence and structure simultaneously, the partially latent flow model jointly generates C_α coordinates and latent variables that can then be put through the decoder. Ensemble generation can be performed by repeated conditional generation with a given sequence. Ensemble-based sequence design can be performed by repeated conditional generation with a given ϕ,ψ-backbone distribution.

of C_α coordinates $\mathbf{x}_i \in \mathbb{R}^{n\times 3}$ and per-residue latent variables $\mathbf{z}_i \in \mathbb{R}^{n\times 8}$, and then decoding to full atomic structure and sequence via a decoder. Generation may be conditioned on any single signal: amino acid sequence $\mathbf{s}$, reference C_α coordinates $\mathbf{x}_{\text{ref}}$, or per-residue ϕ,ψ-backbone distributions $\mathbf{R} \in \mathbb{R}^{n\times 36\times 36}$. A general schematic is shown in Figure 1.

3.2. Preliminary

Following the two-stage partially latent flow matching framework of *La-Proteina* (Geffner et al., 2025), we separate cyclic peptide representation into explicit, cartesian C_α coordinate geometries and learned per-residue latent space representations for residue-type and side-chain coordinates, then perform flow matching in the resulting product space.

Stage 1: Variational autoencoder. An encoder $q_\phi(\mathbf{z}|\mathbf{x},\mathbf{s})$ maps all-atom coordinates in the Atom37 representation ($\mathbb{R}^{n\times 37\times 3}$) and sequence to per-residue latent variables $\mathbf{z} \in \mathbb{R}^{n\times 8}$ via 8 transformer layers with pair-bias attention. The Atom37 representation describes every residue using a fixed 37-atom superset of all standard heavy-atom positions such that each residue can be stored in a single fixed-width tensor. Rich pair features encode inter-residue geometry including asymmetric backbone distances, orientation angles, and sequence separation (see Appendix A for details). The encoder outputs per-residue mean and log-scale for a diagonal Gaussian (Kingma et al., 2014). A decoder $p_\theta(\hat{\mathbf{x}}_{\text{all-atom}}, \hat{\mathbf{s}}|\mathbf{z}, \mathbf{x}_{C_\alpha})$ predicts sequence logits over the 39-type vocabulary from latent variables and C_α positions and reconstructs all-atom coordinates.

Stage 2: Flow matching denoiser. A denoiser D_ψ predicts the velocity field (v-prediction, following *La-Proteina*) that transports Gaussian noise to the data distribution in the joint $(\mathbf{x}_{C_\alpha}, \mathbf{z})$ space.

Per-residue features. The denoiser takes as input the current noisy state at flow time t, C_α coordinates (3d) and latent variables (8d), and predicts velocity fields for both simultaneously. Additional inputs include self-conditioning predictions from the previous integration step (3d + 8d, zeroed on first step) and optional conditioning slots which are each independently zeroed according to the active guidance mode. All inputs are concatenated and linearly projected to the token dimension. Timestep conditioning enters via AdaLN rather than as an input feature.

Sampling. We integrate the learned velocity field over 200 steps using a logarithmic time schedule. Integration uses a hybrid SDE/ODE scheme: for $t < 0.98$, the update combines the learned velocity field with a score-based correction and stochastic noise injection ($\mathrm{d}\mathbf{x} = (\mathbf{v} + g(t)\,\nabla \log p_t)\,\mathrm{d}t + \sqrt{2\,g(t)\,T}\,\mathrm{d}\mathbf{W}$), transitioning to a score-scaled ODE for the final steps ($t > 0.98$). The noise scale parameters (temperatures) control structure/ensemble and sequence diversity. C_α and latent channels use independent schedules (logarithmic for C_α, power for latents) and are integrated jointly. C_α coordinates are re-centered to zero center of mass at each step.

See Appendix A for architectural details.

3.3. D-amino Acid Handling

We define a unified vocabulary of 39 amino acid types: indices 0-19 for the 19 L-amino acids (as well as Glycine which is achiral) and 20-38 for the 19 D-amino acids. All metadata required to featurize L-amino acids is directly applicable to its D-amino counterpart; for example, both L and D will contain the same atom-types and bond angle definitions. Because of this, the Atom37 representation is effortlessly implemented. The 39d sequence feature presented to the denoiser encodes only identity with no explicit chirality-informing features. This enables the model to learn the chirality-dependent backbone preferences without loss of generalization. See Appendix C for more details.

3.4. Backbone Dihedral Distribution Conditioning

We introduce here a per-residue ϕ,ψ-backbone distribution denoiser conditioning that encodes conformational preferences. This conditioning enables ensemble-based sequence design from backbone internal degrees of freedom to sequence. For each sequence in the training set, we compute 2-dimensional per-residue ϕ,ψ-backbone histograms $\mathbf{R}_j \in \mathbb{R}^{36\times 36}$ by binning the ϕ and ψ dihedral angles from all conformers into 10° bins over $[-180^\circ, 180^\circ)$, then normalizing to sum to 1. The flattened histogram $\mathbf{r}_j \in \mathbb{R}^{1296}$ is projected via a learned linear layer:

$$\mathbf{h}_j^{\text{rama}} = W_{\text{rama}}\,\mathbf{r}_j + \mathbf{b}_{\text{rama}}, \quad W_{\text{rama}} \in \mathbb{R}^{128\times 1296} \tag{1}$$

For cyclic peptides, the dihedral angles wrap:

$$\phi_j \text{ uses atoms } (C_{(j-1) \bmod n},\, N_j,\, C_j^\alpha,\, C_j),$$
$$\psi_j \text{ uses atoms } (N_j,\, C_j^\alpha,\, C_j,\, N_{(j+1) \bmod n}).$$

3.5. Multi-Modal Classifier-Free Guidance

We train a single denoiser with three conditioning signals that can be independently toggled:

- **Sequence** ($\mathbf{s}$): one-hot amino acid identity (39d)
- **Coordinates** ($\mathbf{x}_{\text{ref}}$): clean C_α positions (3d)
- **ϕ,ψ-Backbone Distribution** ($\mathbf{R}$): per-residue (ϕ, ψ) histograms (128d projected)

Categorical mode sampling. We refer to these conditioning modes as seq-only, coord-only, and rama-only (named after the Ramachandran plot/diagram used for visualization and representation of the ϕ,ψ-backbone distribution). During training, we sample from four exclusive conditioning modes via a categorical distribution:

$$\begin{aligned} P(\text{seq-only}) &= 0.3 \\ P(\text{coord-only}) &= 0.3 \\ P(\text{rama-only}) &= 0.3 \\ P(\text{unconditional}) &= 0.1 \end{aligned} \tag{2}$$

The three conditioning modes receive equal weight during training to ensure balanced training capacity across signals. The 0.1 unconditional allocation provides exposure to the unconditional mode for effective classifier-free guidance.

At inference, for conditioning signal c, the guided velocity is:

$$\hat{v}(\mathbf{x}_t, t, c) = v_\psi(\mathbf{x}_t, t, \varnothing) + w \cdot \big(v_\psi(\mathbf{x}_t, t, c) - v_\psi(\mathbf{x}_t, t, \varnothing)\big) \tag{3}$$

where w controls guidance strength and $\varnothing$ denotes the unconditional mode.

3.6. Dataset

We curate 6,998 heterochiral cyclic peptide sequences across 7 lengths (6-12 residues, ∼1,000 per length), each with 1,000 conformers generated via Rosetta's generalized kinematic closure protocol (Leman et al., 2020) with full-atom energy minimization, totaling ∼7M structures. This size range was chosen for its potential in therapeutic applications (Ji et al., 2024). A triplet-coverage algorithm was used to generate sequences that contain all 39^3 three-residue motifs according to the full 39-type heterochiral vocabulary. This collection was down-sampled to 1000 sequences per length when building this dataset. All generated conformers include full heavy-atom coordinates. This represents a substantial expansion over prior cyclic peptide ML datasets restricted to shorter lengths and population binning. Appendix F contains Ramachandran plots ((ϕ, ψ)-backbone distribution plots) of residue types and cyclic peptides generated for the dataset.

Data is split randomly at the sequence level: 900 training, 50 validation, and 50 test sequences per length. For constrained peptides, it is well known, and confirmed both experimentally and computationally (Cummings et al., 2019), that deviation by a single residue mutation can drastically change the conformational ensemble; this justifies the random split approach as nearly identical sequences can have very different ensembles. Per-sequence ϕ,ψ-backbone histograms (36×36 per residue) are precomputed from all conformers for each sequence. For more details on the dataset generation see Appendix B.

4. Experiments

4.1. Setup

We perform inference-time parameter tuning on the held-out validation set (see Appendix J) as well as ensemble generation and ensembled-based sequence design on the held-out test set using the three single-signal conditioning modes detailed above (see Appendix H for full results). In all experiments 1000 conformers are generated per sequence (seq-only) or 1000 sequences per ensemble (rama-only). Ensemble reconstruction, sequence recovery, ring closure, and ensemble entropy are all considered through various metrics detailed below. For cluster code cosine similarity, a new data-driven conformational state map is proposed (see Appendix E). Comparison and benchmarking against other structure and sequence predictors (Boltz2 (Passaro et al., 2025) and ProteinMPNN (Dauparas et al., 2022)) are also considered. Finally, the model's ability to learn cyclic peptide symmetries is considered.

4.2. Evaluation Metrics

We evaluate ensemble fidelity using nine complementary metrics that assess distributional similarity, conformational diversity, and sequence prediction accuracy. For each test sequence, the model generates 1,000 conformers; ground truth (GT) consists of up to 1,000 Rosetta-generated conformers.

Ensemble metrics. We assess ensemble fidelity with four complementary measures: **Ramachandran cosine similarity (RamaCos)** compares per-residue (ϕ, ψ) histograms between generated and reference ensembles; **cluster code cosine similarity (ClusterCos)** categorizes each conformer into an string of Greek letters (same length as the peptide) composing 12 data-driven backbone states and compares population vectors (see Appendix E for more details); **PCA sliced Wasserstein distance (PCA-SW)** and **PCA cosine similarity (PCA-CosSim)** project conformers via $\cos / \sin$-transformed dihedral angles into a shared 2D PCA subspace and measure distributional overlap via earth-mover distance and histogram cosine similarity, respectively. Ensemble entropy quantifies conformational diversity as the Shannon entropy over cluster code frequencies.

Sequence metrics. Several sequence metrics are used. **Sequence recovery (SeqRec)** measures the fraction of generated conformers whose autoencoder-decoded sequence exactly matches any ground-truth sequence; **per-residue recovery (PRR)** compares the single most frequent generated sequence against the most frequent ground-truth sequence position-by-position, capturing local accuracy even when exact full-sequence match is combinatorially rare. **Sequence uniqueness (SeqUnique)** counts distinct decoded sequences among 1,000 conformers, diagnosing whether the model has learned a deterministic ensemble-to-sequence mapping (low uniqueness) or lacks sequence information in the conditioning signal (high uniqueness).

Entropy and **ring closure** are also considered. See Appendix D for more details.

4.3. Inference Time Parameter Tuning

We sweep inference-time parameters on a subset of the held-out validation set (10 sequences of length 6 and length 12) under all three conditioning modes. Four axes are explored: guidance weight $w \in \{0.00\text{-}5.00\}$, integration steps $\in \{50\text{-}400\}$, C_α temperature $T_{\text{CA}} \in \{0.05\text{-}1.0\}$, and latent temperature $T_{\text{lat}} \in \{0.05\text{–}1.0\}$. Quality plateaus at 200 integration steps (RamaCos improves from 0.06 at 50 steps to 0.48 at 200, with negligible gains beyond); RamaCos peaks at $w = 0.75$ for all three modes (See Section 5 for discussion on this result); and $T_{\text{CA}} = 0.5$, $T_{\text{lat}} = 0.1$ consistently outperforms other temperature combinations. The two temperature parameters are swept on a grid. Based on these sweeps we define two evaluation configurations: **Config A** (w=0.75, T_{CA}=0.5, T_{lat}=0.1) which we judge to be optimized parameters, and **Config B** (w=1.0, T_{CA}=0.1, T_{lat}=0.1) representing a default set of parameters prior to tuning. Over the test set, Config A uniformly outperforms Config B on all ensemble metrics (e.g. RamaCos 0.524 vs. 0.477 for seq-only; PCA-SW 0.268 vs. 0.649), validating our parameter sweep findings that increasing T_{CA} from 0.1 to 0.5 while reducing w to 0.75 substantially improves ensemble generation without sacrificing accuracy. Full parameter sweep tables are provided in Appendix J.

4.4. Ensemble Generation

We evaluate the conformational ensemble quality on the held-out test set under seq-only conditioning with Config A, generating 1,000 conformers per sequence. The model achieves mean RamaCos of 0.524 (0.570 for 12-mers, 0.403 for 6-mers), indicating strong per-residue distributional agreement with the Rosetta reference ensembles. Figure 2 provides examples of CPGEN generated ensembles against Rosetta ground truth ensembles; it is evident from the plots that RamaCos values of $\sim$0.6 are very strong. PCA-SW is 0.268 overall, with longer peptides showing better distributional overlap (0.178 for 12-mers vs. 0.441 for 6-mers) due to the more disperse, flatter conformational landscape of larger rings. ClusterCos is 0.287 overall and is dominated by short peptides (0.725 for 6-mers) where greater ring strain leads to higher populations of specific, preferred conformers. Generated and ground-truth ensemble entropies are strongly correlated ($R^2 = 0.859$, see Figure 26), confirming that the model reproduces per-sequence conformational diversity rather than collapsing to a single mode or uniformly dispersing. Ring closure is maintained without explicit enforcement: 97.3% of conformers achieve

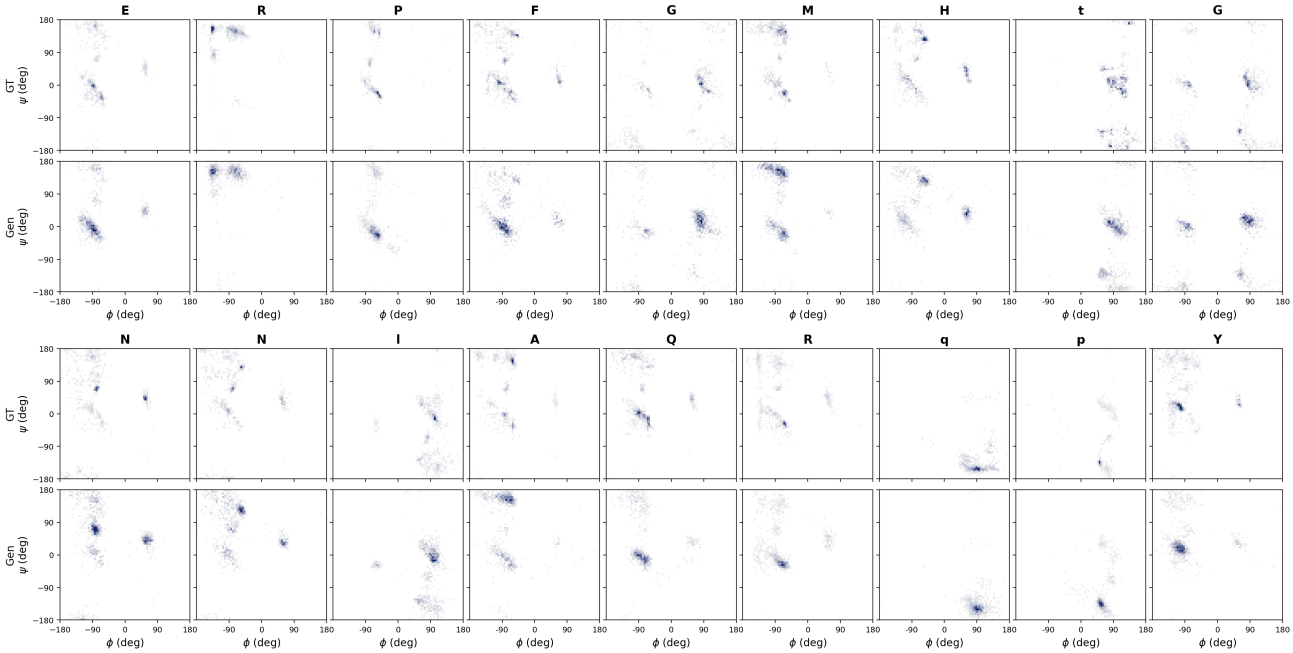

Figure 2. Per-residue Ramachandran plots for the top-2 9-mer ensemble generation test results: ERPFGMHtG (top) and NNlAQRqpY (bottom). Each sequence has Ramachandran plots for the ground truth (GT, Rosetta) on the top and CPGEN generated on the bottom. ERPFGMHtG has a RamaCos of 0.601 and NNlAQRqpY of 0.593

a closing CN distance below 1.5 Å (mean 1.34 Å). For full results over the test set and per length breakdowns see Appendix H; more per-residue Ramachandram plots as well as 2D PCA plots are therein.

4.5. Ensemble-Based Sequence Design

Under rama-only conditioning (Config A), the model infers amino acid sequence from per-residue ϕ, ψ-backbone distributions alone, achieving 60.4% exact sequence recovery and 89.9% PRR overall. Performance scales strongly with peptide length: 12-mers reach 91.3% SeqRec and nearly 100% PRR, while 6-mers achieve only 5.9% SeqRec (exact sequence recovery) and 60.0% PRR with most erroneous predictions being conservative (i.e valine in place of leucine). These results reflect that in shorter rings greater diversity of ϕ,ψ-backbone distributions exist for specific residue-type due to the influence of neighboring residue and ring constraints. Furthermore, SeqUnique is low (12.5 distinct sequences on average per ensemble), indicating the model converges on a consistent set of sequence prediction from the ϕ,ψ-backbone distribution signal. By contrast, coord-only conditioning (see details in Appendix H) achieves 0% SeqRec at all lengths with near-maximal SeqUnique (997 distinct sequences per ensemble), confirming the hypothesis that sequence prediction from C_α coordinates does not uniquely determine the amino acid sequence due to the lack of ensemble information. For full results over the test set and per length breakdowns see Appendix H.

4.6. Comparison with Established Methods

We acknowledge the difficulty in benchmarking our model against other tools for several reasons. First, structure prediction tools are not trained to generate ensembles, but are trained to generate single structures from a specific sequence. Second, attempts to extend structure prediction models to ensemble prediction have not include cyclic peptides. Third, our model and test sets natively include heterochiral peptides that are often excluded from general purpose models for structure prediction and inverse folding. Fourth, inverse folding methods typically operate on a single (or a small number) of structures and not on an ensemble.

Nevertheless, for demonstration purposes we generated a set of 100 L-amino acid containing cyclic peptides (residue lengths 6 and 12), and we performed folding with Boltz2 (Passaro et al., 2025) and inverse folding with ProteinMPNN (Dauparas et al., 2022). It should be noted that Boltz2 does have an explicit cyclic option for structure prediction. Boltz2 was used to generate 1,000 sample conformers of each sequence and ProteinMPNN was used to generate 10 sequences over 100 randomly subsampled peptide backbones from the Rosetta ground truth ensembles. CPGEN was run on seq-only for ensemble generation and rama-only for ensemble-based sequence design. Both baselines operate on the same test sequences under their default settings; CPGEN uses Config A (w=0.75, T_{CA}=0.5, T_{lat}=0.1).

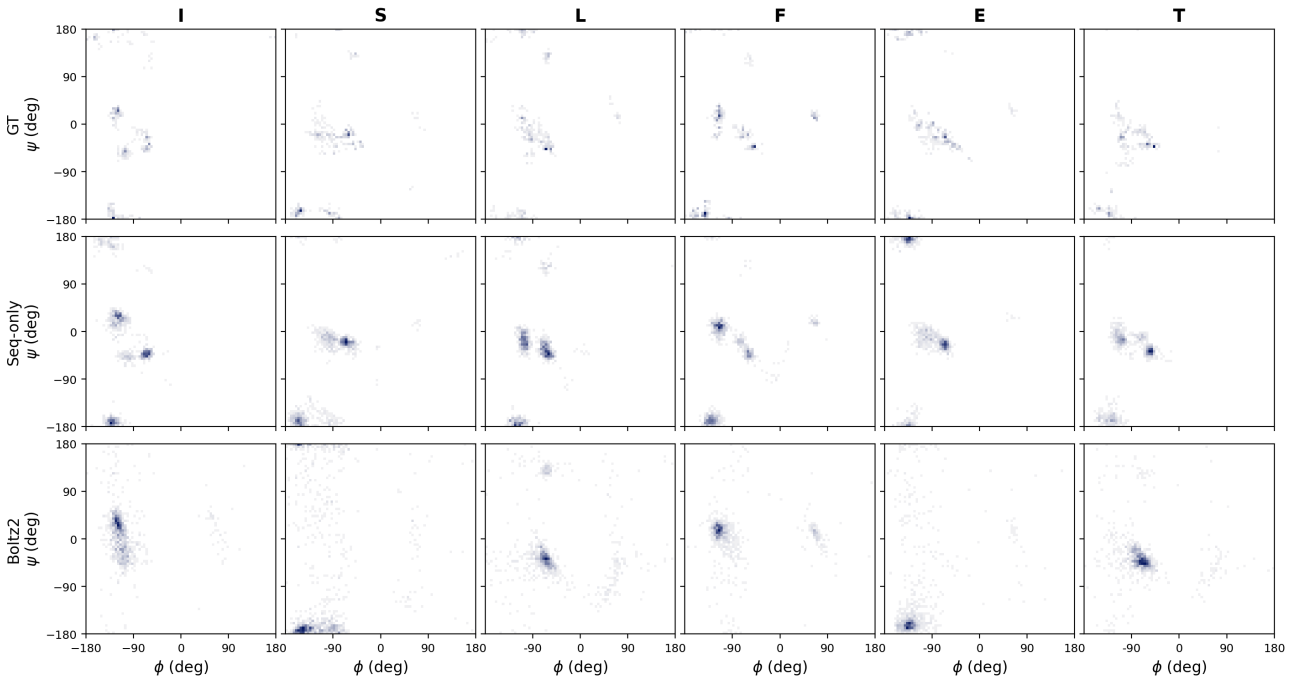

Figure 3. Per-residue Ramachandran plots of the 6-mer (ISLFET) with the strongest metrics for Boltz2. From top to bottom: Rosetta generated ground truth ensemble (GT), CPgen seq-only generate ensemble (seq-only), and the Boltz2 generated ensemble. The Botlz2 RamaCos is 0.265, and the CPgen seq-only RamaCos is 0.444.

Table 1. Ensemble quality: CPGEN (seq-only) vs. Boltz2, both compared to Rosetta ground truth. Mean over 50 sequences per length.

Method	Metric	ALL	6-mers	12-mers
CPGEN	RamaCos ↑	0.509	0.435	0.582
	ClusterCos ↑	0.442	0.882	0.003
	PCA-CosSim ↑	0.379	0.506	0.252
	PCA-SW ↓	0.232	0.302	0.163
Boltz2	RamaCos ↑	0.202	0.159	0.245
	ClusterCos ↑	0.207	0.414	0.000
	PCA-CosSim ↑	0.102	0.170	0.034
	PCA-SW ↓	1.406	1.195	1.617

Table 2. Sequence design: CPGEN (rama-only) vs. ProteinMPNN.

Method	Metric	ALL	6-mers	12-mers
CPGEN	SeqRec ↑	0.454	0.026	0.882
	PRR ↑	0.806	0.617	0.995
	SeqUnique	22.6	40.3	5.0
ProteinMPNN	SeqRec ↑	0.000	0.000	0.000
	PRR ↑	0.121	0.080	0.162
	SeqUnique	597.0	322.7	871.3

Table 1 displays results for the ensemble generation from Boltz2 and the seq-only sampling mode of CPGEN. Across all metrics CPGEN outperforms Boltz2 in generating realistic conformational ensembles. It should be noted from Table 23 that Boltz2 does not display structure prediction collapse, but does in fact generate ensembles of conformers that are diverse but with populations that simply do not recapitulate the ground truth ensemble. Figure 3 shows the ground truth, CPGEN, and Boltz2 predicted ensembles. Table 2 shows the predictive power of CPGEN's rama-only sampling to predict cyclic peptide sequences. The current industry standard, ProteinMPNN, is unable to recover the exact sequence at any length, generates a large number of unique sequences (very high noise predictions), and fails at PRR as well. Conversely, CPGEN has much strong results across every metric and peptide length. Sequence recovery of cyclic hexapeptides is challenging for CPGEN as well (see below for discussion). Additional details including error bars for these values can be found in Appendix I.

4.7. Learned Symmetries: Chiral Inversion and Cyclic Permutation

Cyclic peptides possess two inherent symmetries: (1) the ring has no fixed start residue, so all cyclic permutations of a sequence represent the same molecule, and (2) inverting all chirality (L→D, D→L) produces a mirror-image peptide whose backbone dihedral distributions should satisfy $(\phi, \psi) \rightarrow (-\phi, -\psi)$. We test whether the model has learned these symmetries without explicit equivariance constraints by generating 1,000-conformer ensembles for all 6 cyclic permutations and their 6 chiral inversions of 5 hexapeptide

(6-mers) test sequences, then comparing per-residue ϕ,ψ-backbone distributions with appropriate alignment (cyclic index shift for permutations; (ϕ, ψ) inversion for chiral inversions).

Table 3. Symmetry test results. PermCos: mean RamaCos across cyclic shifts 1-5. ChiralCos: mirrored RamaCos of chiral inversion at shift=0. ChPermCos: mean mirrored RamaCos across all 6 chiral inverted permutations.

Sequence	PermCos ↑	ChiralCos ↑	ChPermCos ↑
TanplL	0.924	0.820	0.817
NvIrST	0.881	0.831	0.810
LdTcfq	0.835	0.789	0.777
sLkYQF	0.838	0.806	0.779
WAsQHR	0.829	0.770	0.739
Mean	**0.861**	**0.803**	**0.784**

The model achieves mean aligned RamaCos of 0.861 across cyclic permutations, indicating that it produces highly similar conformational ensembles regardless of sequence starting position. For chiral inversion, mirrored RamaCos averages 0.803, confirming that the model has learned that D-amino acids produce approximately mirror-image backbone distributions relative to their L-counterparts. Both symmetries emerge purely from the training data without any architectural enforcement (no cyclic positional encoding, no chiral augmentation), demonstrating that the flow matching denoiser implicitly learns the symmetry structure of cyclic peptides.

5. Discussion

ϕ,ψ-backbone distribution as sequence fingerprints. The 0% SeqRec of coord-only sampling (see Table 12) over the test set and the ProteinMPNN benchmarking confirm that conformational ensemble *distributions* carry amino acid identity information that single static structures do not. The fundamental advantage here is that ϕ,ψ-backbone *distributions* encode coupled residue-level, population-weighted conformational preferences of each amino acid in a given cyclic peptide, whereas point coordinates carry no such information. This positions ϕ,ψ-backbone distribution conditioned generation as a novel ensemble-based approach to sequence design that complements the single-structure inverse folding methods suitable for more stable molecular modalities.

Guidance below 1.0 generates better ensembles. It was discovered during parameter tuning that the model experiences some structure prediction mode collapse during seq-only sampling which can be recovered by tuning guidance down rather than up. This allows for the model to generate more diffuse ensembles with more diverse structures that capture the true ensemble statistics better. The optimal guidance parameter ($w = 0.75$) contrasts with $w > 1$ in image generation (Ho & Salimans, 2022). This finding leads us to believe that when the target is a distribution generated by repeated sampling that excessive guidance concentrates probability on dominant conformations at the expense of less populated ones that are needed for accurate distributional generation. This guidance/diversity trade-off suggests that distributional generation requires different tuning heuristics than single-output settings.

Length-dependent Results. Cyclic peptides possess a non-trivial length to ensemble relationship. At shorter lengths, closer to six residues, cyclic peptides take on fewer conformations due to ring strain (lower ring flexibility). It is well known that small changes in sequence can lead to very different ensembles. At larger lengths, around twelve residues, less ring strain leads to much more flexible molecules with much smoother energy landscapes and many thermally accessible states. Finally, at much larger sizes (not in this study), cyclic peptides begin to adopt folds similar to linear peptides that actually stabilize the conformational landscape. In our study, this is reflected in the ClusterCos metric being easier to predict for smaller peptides but more difficult for larger peptides as the number of accessible conformers grows extremely large. Conversely, the RamaCos and SeqRec metrics are trivial for larger peptides where flexibility allows the individual residues to statistically orient themselves independently and revert to their mean behavior.

Well-Structuredness and Entropy Predictions. An extremely difficult and important task in cyclic peptide ensemble prediction and sequence design is the determination of well-structuredness. That is, the prediction of cyclic peptides with a single or a few dominant conformations. Of the four most well structured peptides in the test dataset, CPGEN generates highly accurate conformational ensembles for each (see Appendix H.3.1). Furthermore, CPGEN generates ensembles that have similar entropies as those of ground truth ensembles (see Appendix H.3). Figure 4 shows an example of hexapeptide with Ramachandran plots and representative structures for specific cluster code conformations as well as a test set hexapeptide ensemble entropy parity plot of Rosetta ground truth versus CPGEN.

Speed Advantage. At batch size $1{,}000$, CPGEN generates a complete conformational ensemble in 12 seconds for 6-mers and 25 seconds for 12-mers on a single GPU, an 80–$400\times$ speedup over Rosetta GenKIC on a single CPU core ($\sim$2.3 hours and $\sim$1.6 hours per sequence, respectively). The entire $\sim$7 million-conformer training dataset, which required $9{,}400$ CPU-hours with Rosetta, could be regenerated in $\sim$24 GPU-hours. This throughput enables rapid screening of large sequence libraries at ensemble resolution,

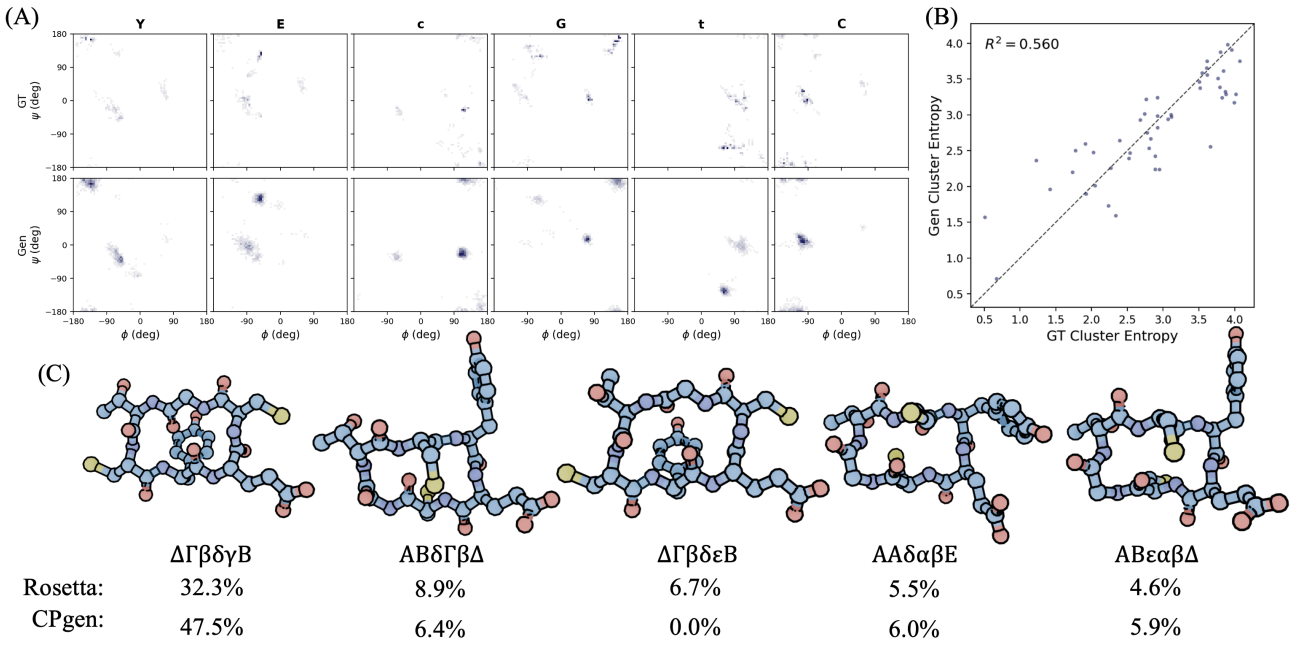

Figure 4. (A) Sequence YEcGtC Ramachandran plots for the Rosetta generated ground truth (GT) ensemble and the CPGEN seq-only generated ensemble. The RamaCos value is 0.439 showing strong agreement across the entire sequence. (B) The YEcGtC conformational ensemble has a ground truth cluster code entropy of 2.89 (2.24 for the CPGEN seq-only generated). A parity plot of ground truth entropy versus entropy of CPGEN generated ensembles is displayed (top-right) showing good correlation. (C) Five representative structures with their associated cluster codes, ground truth populations, and CPGEN seq-only generated populations. CPGEN does very well capturing the overall populations and only misses on a single cluster code that is only one digit away from the highest population code. This missing population in the CPGEN generated ensemble seems to be absorbed by the dominant conformation.

a regime previously inaccessible to physics-based sampling. For a full speed comparison see Appendix G.

Limitations. Ground truth ensembles are Rosetta sampled, carrying force field biases; validation against NMR observables is an important next step. The 39-type amino acids vocabulary does not include other non-canonical residue types such as N-methylated and β-amino acids that are common in therapeutics discovery. Extension to longer macrocycles and non-peptidic linkers is ongoing.

6. Conclusion

CPGEN adapts the partially latent flow matching of *La-Proteina* for heterochiral cyclic peptide ensemble generation, demonstrating three capabilities: fast sequence-conditioned ensemble generation, ensemble-based sequence design from ϕ,ψ-backbone distributions, and native D-amino acid support. Other findings include: (1) conformational ensemble descriptors carry sufficient information for sequence generation, suggesting a new paradigm of ensemble-based sequence design, (2) guidance of $w < 1$ may improve distributional generation in the peptide ensemble generation context, and (3) the speed advantage of generative machine learning based approaches to ensemble generation over physics-based methods is further evidenced.

Impact Statement

This paper presents work whose goal is to advance computational methods for therapeutic peptide design. Generative models for peptide conformer ensembles could accelerate drug discovery by reducing the computational cost of conformational analysis. We do not foresee specific negative societal consequences beyond those common to drug design tools.

LLM usage

LLMs were used during the preparation of the manuscript solely as writing aids to catch typographical and grammatical errors and to improve writing style and latex formatting.

Code availability

Removed for double-blind review.

Acknowledgments

Anonymized for double-blind review.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016): 493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w.

Abrudan, A., Ojeda, S. P., Joshi, C. K., Greenig, M., Engelberger, F., Khmelinskaia, A., Meiler, J., Vendruscolo, M., and Knowles, T. P. J. Multi-state Protein Design with DynamicMPNN, October 2025. URL http://arxiv.org/abs/2507.21938. arXiv:2507.21938 [cs].

A. Viarengo-Baker, L., E. Brown, L., A. Rzepiela, A., and Whitty, A. Defining and navigating macrocycle chemical space. *Chemical Science*, 12(12): 4309–4328, 2021. doi: 10.1039/D0SC05788F. URL https://pubs.rsc.org/en/content/articlelanding/2021/sc/d0sc05788f.

Bayaraa, N., Secor, M., Descoteaux, M. L., and Lin, Y.-S. Fast Generation of Simulation-Quality Structural Ensembles of Mixed-Chirality Cyclic Peptides via Diffusion Models. *Journal of Chemical Theory and Computation*, 22(6):3103–3113, March 2026. ISSN 1549-9618. doi: 10.1021/acs.jctc.5c01862. URL https://doi.org/10.1021/acs.jctc.5c01862.

Bhardwaj, G., Mulligan, V. K., Bahl, C. D., Gilmore, J. M., Harvey, P. J., Cheneval, O., Buchko, G. W., Pulavarti, S. V. S. R. K., Kaas, Q., Eletsky, A., Huang, P.-S., Johnsen, W. A., Greisen, P. J., Rocklin, G. J., Song, Y., Linsky, T. W., Watkins, A., Rettie, S. A., Xu, X., Carter, L. P., Bonneau, R., Olson, J. M., Coutsias, E., Correnti, C. E., Szyperski, T., Craik, D. J., and Baker, D. Accurate de novo design of hyperstable constrained peptides. *Nature*, 538(7625):329–335, October 2016. ISSN 1476-4687. doi: 10.1038/nature19791. URL https://www.nature.com/articles/nature19791.

Bose, A. J., Akhound-Sadegh, T., Huguet, G., Fatras, K., Rector-Brooks, J., Liu, C.-H., Nica, A. C., Korablyov, M., Bronstein, M., and Tong, A. SE(3)-Stochastic Flow Matching for Protein Backbone Generation, April 2024. URL http://arxiv.org/abs/2310.02391. arXiv:2310.02391 [cs].

Calvo-Barreiro, L., Secor, M., Damjanovic, J., Abdel-Rahman, S. A., Lin, Y.-S., and Gabr, M. Computational Design of a Bicyclic Peptide Inhibitor Targeting the ICOS/ICOS-L Protein–Protein Interaction. *Chemical Biology & Drug Design*, 105(5):e70117, 2025. ISSN 1747-0285. doi: 10.1111/cbdd.70117. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cbdd.70117. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cbdd.70117.

Cummings, A. E., Miao, J., Slough, D. P., McHugh, S. M., Kritzer, J. A., and Lin, Y.-S. -Branched Amino Acids Stabilize Specific Conformations of Cyclic Hexapeptides. *Biophysical Journal*, 116(3):433–444, February 2019. ISSN 0006-3495. doi: 10.1016/j.bpj.2018.12.015. URL https://www.sciencedirect.com/science/article/pii/S0006349518345296.

Damjanovic, J., Miao, J., Huang, H., and Lin, Y.-S. Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. *Chemical Reviews*, 121(4):2292–2324, February 2021. ISSN 0009-2665. doi: 10.1021/acs.chemrev.0c01087. URL https://doi.org/10.1021/acs.chemrev.0c01087.

Danelius, E., Poongavanam, V., Peintner, S., Wieske, L. H. E., Erdélyi, M., and Kihlberg, J. Solution Conformations Explain the Chameleonic Behaviour of Macrocyclic Drugs. *Chemistry – A European Journal*, 26(23):5231–5244, 2020. ISSN 1521-3765. doi: 10.1002/chem.201905599. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.201905599. _eprint: https://chemistry-europe.onlinelibrary.wiley.com/doi/pdf/10.1002/chem.201905599.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A. K., King, N. P., and Baker, D. Robust deep learning–based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, October 2022. doi: 10.1126/science.add2187. URL https://www.science.org/doi/10.1126/science.add2187.

Dauparas, J., Lee, G. R., Pecoraro, R., An, L., Anishchenko, I., Glasscock, C., and Baker, D. Atomic context-conditioned protein sequence design using LigandMPNN. *Nature Methods*, 22(4):717–723, April 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02626-1. URL https://www.nature.com/articles/s41592-025-02626-1.

Feller, A. L. and Wilke, C. O. Peptide-specific chemical language model successfully predicts membrane diffusion of cyclic peptides, August 2024. URL https://www.biorxiv.org/content/10.1101/2024.08.09.607221v1. Pages: 2024.08.09.607221 Section: New Results.

Feller, A. L., Secor, M., Swanson, S., Wilke, C. O., and Deibler, K. Scaling SMILES-based chemical language models for therapeutic peptide engineering, April 2026. URL https://www.biorxiv.org/content/10.64898/2026.01.06.697994v3. ISSN: 2692-8205 Pages: 2026.01.06.697994 Section: New Results.

Gao, Z., Tan, C., Chacon, P., and Li, S. Z. PIFOLD: TOWARD EFFECTIVE AND EFFICIENT PROTEIN INVERSE FOLDING. 2023.

Ge, Y., Zhang, S., Erdelyi, M., and Voelz, V. A. Solution-State Preorganization of Cyclic -Hairpin Ligands Determines Binding Mechanism and Affinities for MDM2. *Journal of Chemical Information and Modeling*, 61 (5):2353–2367, May 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c00029. URL https://doi.org/10.1021/acs.jcim.1c00029.

Geffner, T., Didi, K., Cao, Z., Reidenbach, D., Zhang, Z., Dallago, C., Kucukbenli, E., Kreis, K., and Vahdat, A. La-Proteina: Atomistic Protein Generation via Partially Latent Flow Matching, July 2025. URL http://arxiv.org/abs/2507.09466. arXiv:2507.09466 [cs].

Grambow, C. A., Weir, H., Diamant, N. L., Scalia, G., Biancalani, T., and Chuang, K. V. Accurate and Efficient Structural Ensemble Generation of Macrocyclic Peptides using Internal Coordinate Diffusion, August 2024. URL http://arxiv.org/abs/2305.19800. arXiv:2305.19800 [q-bio].

Ho, J. and Salimans, T. Classifier-Free Diffusion Guidance, July 2022. URL http://arxiv.org/abs/2207.12598. arXiv:2207.12598 [cs].

Hosono, Y., Uchida, S., Shinkai, M., Townsend, C. E., Kelly, C. N., Naylor, M. R., Lee, H.-W., Kanamitsu, K., Ishii, M., Ueki, R., Ueda, T., Takeuchi, K., Sugita, M., Akiyama, Y., Lokey, S. R., Morimoto, J., and Sando, S. Amide-to-ester substitution as a stable alternative to N-methylation for increasing membrane permeability in cyclic peptides. *Nature Communications*, 14(1): 1416, March 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-36978-z. URL https://www.nature.com/articles/s41467-023-36978-z.

Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., and Rives, A. Learning inverse folding from millions of predicted structures. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 8946–8970. PMLR, June 2022. URL https://proceedings.mlr.press/v162/hsu22a.html.

Hui, T., Secor, M., Ho, M. N., Bayaraa, N., and Lin, Y.-S. Molecular Dynamics (MD)-Derived Features for Canonical and Noncanonical Amino Acids. *Journal of Chemical Information and Modeling*, 65(4):1837–1849, February 2025. ISSN 1549-9596. doi: 10.1021/acs.jcim.4c02102. URL https://doi.org/10.1021/acs.jcim.4c02102.

Ingraham, J. B., Baranov, M., Costello, Z., Barber, K. W., Wang, W., Ismail, A., Frappier, V., Lord, D. M., Ng-Thow-Hing, C., Van Vlack, E. R., Tie, S., Xue, V., Cowles, S. C., Leung, A., Rodrigues, J. V., Morales-Perez, C. L., Ayoub, A. M., Green, R., Puentes, K., Oplinger, F., Panwar, N. V., Obermeyer, F., Root, A. R., Beam, A. L., Poelwijk, F. J., and Grigoryan, G. Illuminating protein space with a programmable generative model. *Nature*, 623(7989):1070–1078, November 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06728-8. URL https://www.nature.com/articles/s41586-023-06728-8.

Ji, X., Nielsen, A. L., and Heinis, C. Cyclic Peptides for Drug Development. *Angewandte Chemie*, 136(3):e202308251, 2024. ISSN 1521-3757. doi: 10.1002/ange.202308251. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ange.202308251. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ange.202308251.

Jing, B., Berger, B., and Jaakkola, T. AlphaFold Meets Flow Matching for Generating Protein Ensembles, September 2024. URL http://arxiv.org/abs/2402.04845. arXiv:2402.04845 [q-bio].

Kingma, D. P., Rezende, D. J., Mohamed, S., and Welling, M. Semi-Supervised Learning with Deep Generative Models, October 2014. URL http://arxiv.org/abs/1406.5298. arXiv:1406.5298 [cs].

Koch, G., Engstrom, A., Taechalertpaisarn, J., Faris, J., Ono, S., Naylor, M. R., and Lokey, R. S. Chromatographic Determination of Permeability-Relevant Lipophilicity Facilitates Rapid Analysis of Macrocyclic Peptide Scaffolds. *Journal of Medicinal Chemistry*, 67(21):19612–19622, November 2024. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.4c01956. URL https://doi.org/10.1021/acs.jmedchem.4c01956.

Krueger, R. K., Brenner, M. P., and Shrinivas, K. Generalized design of sequence–ensemble–function relationships for intrinsically disordered proteins.

Nature Computational Science, pp. 1–12, October 2025. ISSN 2662-8457. doi: 10.1038/s43588-025-00881-y. URL https://www.nature.com/articles/s43588-025-00881-y.

Laio, A. and Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99 (20):12562–12566, October 2002. doi: 10.1073/pnas.202427399. URL https://www.pnas.org/doi/10.1073/pnas.202427399.

Leman, J. K., Weitzner, B. D., Lewis, S. M., Adolf-Bryfogle, J., Alam, N., Alford, R. F., Aprahamian, M., Baker, D., Barlow, K. A., Barth, P., Basanta, B., Bender, B. J., Blacklock, K., Bonet, J., Boyken, S. E., Bradley, P., Bystroff, C., Conway, P., Cooper, S., Correia, B. E., Coventry, B., Das, R., De Jong, R. M., DiMaio, F., Dsilva, L., Dunbrack, R., Ford, A. S., Frenz, B., Fu, D. Y., Geniesse, C., Goldschmidt, L., Gowthaman, R., Gray, J. J., Gront, D., Guffy, S., Horowitz, S., Huang, P.-S., Huber, T., Jacobs, T. M., Jeliazkov, J. R., Johnson, D. K., Kappel, K., Karanicolas, J., Khakzad, H., Khar, K. R., Khare, S. D., Khatib, F., Khramushin, A., King, I. C., Kleffner, R., Koepnick, B., Kortemme, T., Kuenze, G., Kuhlman, B., Kuroda, D., Labonte, J. W., Lai, J. K., Lapidoth, G., Leaver-Fay, A., Lindert, S., Linsky, T., London, N., Lubin, J. H., Lyskov, S., Maguire, J., Malmström, L., Marcos, E., Marcu, O., Marze, N. A., Meiler, J., Moretti, R., Mulligan, V. K., Nerli, S., Norn, C., Ó'Conchúir, S., Ollikainen, N., Ovchinnikov, S., Pacella, M. S., Pan, X., Park, H., Pavlovicz, R. E., Pethe, M., Pierce, B. G., Pilla, K. B., Raveh, B., Renfrew, P. D., Burman, S. S. R., Rubenstein, A., Sauer, M. F., Scheck, A., Schief, W., Schueler-Furman, O., Sedan, Y., Sevy, A. M., Sgourakis, N. G., Shi, L., Siegel, J. B., Silva, D.-A., Smith, S., Song, Y., Stein, A., Szegedy, M., Teets, F. D., Thyme, S. B., Wang, R. Y.-R., Watkins, A., Zimmerman, L., and Bonneau, R. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nature Methods*, 17(7): 665–680, July 2020. ISSN 1548-7105. doi: 10.1038/s41592-020-0848-2. URL https://www.nature.com/articles/s41592-020-0848-2.

Lewis, S., Hempel, T., Jiménez-Luna, J., Gastegger, M., Xie, Y., Foong, A. Y. K., Satorras, V. G., Abdin, O., Veeling, B. S., Zaporozhets, I., Chen, Y., Yang, S., Schneuing, A., Nigam, J., Barbero, F., Stimper, V., Campbell, A., Yim, J., Lienen, M., Shi, Y., Zheng, S., Schulz, H., Munir, U., Clementi, C., and Noé, F. Scalable emulation of protein equilibrium ensembles with generative deep learning, December 2024. URL https://www.biorxiv.org/content/10.1101/2024.12.05.626885v1. Pages: 2024.12.05.626885 Section: New Results.

Li, Q., Daumiller, D., and Bryant, P. RareFold: Structure prediction and design of proteins with noncanonical amino acids, May 2025. URL https://www.biorxiv.org/content/10.1101/2025.05.19.654846v1. Pages: 2025.05.19.654846 Section: New Results.

Linker, S. M., Schellhaas, C., Kamenik, A. S., Veldhuizen, M. M., Waibl, F., Roth, H.-J., Fouché, M., Rodde, S., and Riniker, S. Lessons for Oral Bioavailability: How Conformationally Flexible Cyclic Peptides Enter and Cross Lipid Membranes. *Journal of Medicinal Chemistry*, 66 (4):2773–2788, February 2023. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.2c01837. URL https://doi.org/10.1021/acs.jmedchem.2c01837.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow Matching for Generative Modeling, October 2022. URL https://arxiv.org/abs/2210.02747v2.

Mao, Q., Shang, T., Xu, W., Zhai, S., Zhang, C., Guo, J., Su, A., Li, C., and Duan, H. NCPepFold: Accurate Prediction of Noncanonical Cyclic Peptide Structures via Cyclization Optimization with Multigranular Representation. *Journal of Chemical Theory and Computation*, 21(9):4979–4991, May 2025. ISSN 1549-9618. doi: 10.1021/acs.jctc.5c00139. URL https://doi.org/10.1021/acs.jctc.5c00139.

Miao, J., Descoteaux, M. L., and Lin, Y.-S. Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chemical Science*, 12(44):14927–14936, November 2021. ISSN 2041-6539. doi: 10.1039/D1SC05562C. URL https://pubs.rsc.org/en/content/articlelanding/2021/sc/d1sc05562c.

Muttenthaler, M., King, G. F., Adams, D. J., and Alewood, P. F. Trends in peptide drug discovery. *Nature Reviews Drug Discovery*, 20(4):309–325, April 2021. ISSN 1474-1784. doi: 10.1038/s41573-020-00135-8. URL https://www.nature.com/articles/s41573-020-00135-8.

Noé, F., Olsson, S., Köhler, J., and Wu, H. Boltzmann Generators – Sampling Equilibrium States of Many-Body Systems with Deep Learning, July 2019. URL http://arxiv.org/abs/1812.01729. arXiv:1812.01729 [stat].

Ono, S., Naylor, M. R., Townsend, C. E., Okumura, C., Okada, O., Lee, H.-W., and Lokey, R. S. Cyclosporin A: Conformational Complexity and Chameleonicity. *Journal of Chemical Information and Modeling*, 61(11): 5601–5613, November 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c00771. URL https://doi.org/10.1021/acs.jcim.1c00771.

Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V. R., Getz, N., Portnoi, T., Roy, J., Stark, H., Kwabi-Addo, D., Beaini, D., Jaakkola, T., and Barzilay, R. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction, June 2025. URL https://www.biorxiv.org/content/10.1101/2025.06.14.659707v1. Pages: 2025.06.14.659707 Section: New Results.

Poongavanam, V., Wieske, L. H. E., Peintner, S., Erdélyi, M., and Kihlberg, J. Molecular chameleons in drug discovery. *Nature Reviews Chemistry*, 8(1):45–60, January 2024. ISSN 2397-3358. doi: 10.1038/s41570-023-00563-1. URL https://www.nature.com/articles/s41570-023-00563-1.

Wang, L., Wang, N., Zhang, W., Cheng, X., Yan, Z., Shao, G., Wang, X., Wang, R., and Fu, C. Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1):48, February 2022. ISSN 2059-3635. doi: 10.1038/s41392-022-00904-4. URL https://www.nature.com/articles/s41392-022-00904-4.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., Lauko, A., De Bortoli, V., Mathieu, E., Ovchinnikov, S., Barzilay, R., Jaakkola, T. S., DiMaio, F., Baek, M., and Baker, D. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, August 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06415-8. URL https://www.nature.com/articles/s41586-023-06415-8.

White, T. R., Renzelman, C. M., Rand, A. C., Rezai, T., McEwen, C. M., Gelev, V. M., Turner, R. A., Linington, R. G., Leung, S. S. F., Kalgutkar, A. S., Bauman, J. N., Zhang, Y., Liras, S., Price, D. A., Mathiowetz, A. M., Jacobson, M. P., and Lokey, R. S. On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nature Chemical Biology*, 7(11):810–817, November 2011. ISSN 1552-4469. doi: 10.1038/nchembio.664. URL https://www.nature.com/articles/nchembio.664.

Xiao, W., Jiang, W., Chen, Z., Huang, Y., Mao, J., Zheng, W., Hu, Y., and Shi, J. Advance in peptide-based drug development: delivery platforms, therapeutics and vaccines. *Signal Transduction and Targeted Therapy*, 10(1):74, March 2025. ISSN 2059-3635. doi: 10.1038/s41392-024-02107-5. URL https://www.nature.com/articles/s41392-024-02107-5.

Yu, Z., Liu, Y., Lin, G., Jiang, W., and Chen, M. ESMAdam: a plug-and-play all-purpose protein ensemble generator, January 2025. URL https://www.biorxiv.org/content/10.1101/2025.01.19.633818v1. Pages: 2025.01.19.633818 Section: New Results.

Zhang, C., Wang, W., Zhu, N., Cao, Z., Wu, Y., Mao, Q., Zhu, C., Zhang, C., Guo, J., and Duan, H. AlphaFold3 for Noncanonical Cyclic Peptide Modeling: Hierarchical Benchmarking Reveals Accuracy and Practical Guidelines. *Journal of Chemical Information and Modeling*, 65(18):9777–9789, September 2025. ISSN 1549-9596. doi: 10.1021/acs.jcim.5c01393. URL https://doi.org/10.1021/acs.jcim.5c01393.

A. Architecture and Training

The architecture follows the partially latent design of *La-Proteina*, adapted for a cyclic peptide setting with a 39 residue-type heterochiral vocabulary and multi-modal CFG conditioning.

A.1. Variational Autoencoder

Encoder. The encoder maps all-atom coordinates $\mathbf{x} \in \mathbb{R}^{n\times 37\times 3}$ and residue type to per-residue latent variables via 8 transformer layers ($d_{\text{token}} = 512$, $d_{\text{pair}} = 128$, $d_{\text{cond}} = 128$, 12 attention heads). Input token features are:

- One-hot residue type (39d)
- All-atom absolute coordinates, flattened ($37 \times 3 = 111$d) with 37d atom mask
- All-atom coordinates relative to C_α ($37 \times 3 = 111$d) with 37d atom mask
- Binned backbone dihedral angles (ϕ, ψ, ω): 3 angles $\times$ 21 bins $= 63$d
- Binned sidechain torsion angles (χ_1–χ_4): 4 angles $\times$ 21 bins $+4$-dim validity mask $= 88$d
- Chain break indicator (1d) and chain index (1d)

Pair features consist of:

- Relative sequence separation encoding (127d)
- Asymmetric backbone pair distances $\text{CA}_i \to \{N, \text{CA}, C, \text{CB}\}_j$, binned ($4 \times 21 = 84$d)
- Pairwise residue orientations: 5 inter-residue angles (θ_{12}, θ_{21}, ϕ_{12}, ϕ_{21}, ω) using N, CA, CB atoms, binned ($5 \times 21 = 105$d)
- Same/different chain indicator (1d)

The encoder output is projected via LayerNorm $\to$ Linear to $2 \times 8 = 16$d per residue, split into mean $\boldsymbol{\mu}$ and $\log \boldsymbol{\sigma}$. During training, $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \cdot \exp(\log \boldsymbol{\sigma})$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$; at inference, $\mathbf{z} = \boldsymbol{\mu}$.

Decoder. The decoder takes C_α coordinates and latent variables $(\mathbf{x}_{\text{CA}}, \mathbf{z})$ as input through 8 transformer layers ($d_{\text{token}} = 512$, $d_{\text{pair}} = 128$, 12 heads) with the same pair-bias attention architecture. Token features: sinusoidal position embedding (256d), C_α coordinates (3d), latent space (8d). Pair features: sequence separation (127d) + binned C_α pairwise distances (30 bins, 0.1–3.0 nm).

Two output heads project via LayerNorm $\to$ Linear:

- **Coordinate head**: predicts $37 \times 3 = 111$d per residue as offsets relative to the input C_α position. The C_α slot is zeroed before adding the input C_α , ensuring the output C_α exactly matches the input.
- **Sequence head**: predicts logits over the 39-type amino acid vocabulary per residue.

Training. The autoencoder is trained with three loss terms: L2 coordinate reconstruction loss on all atoms, cross-entropy sequence prediction loss, and KL divergence regularization. KL is annealed linearly over the first $50{,}000$ steps to prevent posterior collapse. Optimizer: AdamW with AMSGrad, learning rate 10^{-4}. Training uses 6 nodes $\times$ 8 GPUs with distributed data parallel, batch size 4 per GPU. EMA with decay 0.999 is applied for evaluation. See more details below.

A.2. Flow Matching Denoiser

Architecture. The denoiser uses 14 transformer layers ($d_{\text{token}} = 768$, $d_{\text{pair}} = 256$, $d_{\text{cond}} = 256$, 12 heads) with pair-bias attention, adaptive layer normalization (AdaLN) for timestep conditioning, QK-LayerNorm for training stability, gated attention output, and SwiGLU feedforward layers with $8\times$ inner expansion (halved to $4\times$ after gating) . Adaptive output scaling (sigmoid-gated, zero-initialized) controls each layer's residual contribution.

Token features.

- Noisy C_α coordinates (3d)
- Per-residue latent space (8d)
- Self-conditioning (Chen et al. 2023 (self-conditioning)): previous step's C_α prediction (3d) + latent prediction (8d), concatenated (zeroed on first step and with 50% probability during training)
- Optional sequence one-hot (39d, zeroed in non-seq modes)
- Optional clean C_α coordinates (3d, zeroed in non-coords modes)
- Optional ϕ,ψ-backbone distributions histogram projection ($1296 \to 128$d via learned linear layer, zeroed in non-rama modes)

Pair features.

- Sequence separation encoding (127d)
- Binned noisy C_α pairwise distances (30 bins)
- Self-conditioning distances: binned distances from previous C_α prediction (30 bins)
- Optional binned clean C_α distances (30 bins, zeroed in non-coords modes)

Pair features are projected to d_{pair} via Linear $\to$ LayerNorm, then conditioned on timestep via pair-level AdaLN. The pair time conditioning takes concatenated time embeddings (t_{CA}, t_{lat}) projected to d_{cond}.

Conditioning. The conditioning vector is built from two independent time embeddings (DDPM-style sinusoidal, 256d each): one for the C_α channel and one for the latent channel. These are concatenated, projected, and passed through two SwiGLU transition layers to produce the per-residue conditioning vector $\mathbf{c} \in \mathbb{R}^{n \times 256}$ used by AdaLN in every transformer block.

Output heads. Two LayerNorm $\to$ Linear heads predict per-residue v-prediction targets: C_α velocity (3d) and latent velocity (8d).

A.3. Transformer Block

Each transformer block consists of:

1. AdaLN pre-normalization conditioned on $\mathbf{c}$: $\hat{\mathbf{x}} = \sigma(W_\gamma \,\text{LN}_c(\mathbf{c})) \cdot \text{LN}(\mathbf{x}) + W_\beta \,\text{LN}_c(\mathbf{c})$, where LN_c is a learned LayerNorm on the conditioning
2. Pair-bias multi-head attention with gating: bias $\mathbf{b}_{ij} = W_b \cdot \text{LN}(\text{pair}_{ij})$ added to attention logits, output gated by $\sigma(W_g \mathbf{x})$. QK-LayerNorm stabilizes attention magnitude.
3. Adaptive output scaling (sigmoid, zero-init) on the attention residual
4. SwiGLU transition: $\text{SwiGLU}(\mathbf{x}) = (\text{SiLU}(\mathbf{g}) \odot \mathbf{x})$ with $8\times$ inner expansion (split to $4\times$ after gating), pre-normalized by AdaLN, post-scaled by adaptive output scaling

A.4. Flow Matching

Interpolation. We use the linear interpolation path of conditional flow matching (Lipman et al. 2023, ICLR (Flow Matching)):

$$\mathbf{x}_t = (1-t)\,\mathbf{x}_0 + t\,\mathbf{x}_1, \quad t \in [0,1] \tag{4}$$

where $\mathbf{x}_0 \sim \mathcal{N}(0, I)$ is noise and $\mathbf{x}_1$ is the data.

Loss. With velocity parameterization (v-prediction), the network predicts the velocity field $v_\psi(\mathbf{x}_t, t)$. The training loss converts to an $\mathbf{x}_1$-prediction loss:

$$\mathcal{L}_{\text{FM}} = \frac{1}{(1-t)^2} \frac{1}{n} \sum_{j=1}^{n} \|\mathbf{x}_{1,j} - (\mathbf{x}_{t,j} + (1-t)\, v_\psi(\mathbf{x}_t, t)_j)\|^2 \tag{5}$$

The $(1-t)^{-2}$ weighting upweights loss near $t = 1$ (close to data), following *La-Proteina*.

Time sampling. Training time is drawn from a Beta-uniform mixture: $t \sim (1-p_u)\,\text{Beta}(\alpha, \beta) + p_u\,\text{Uniform}(0, 1)$ with default parameters $\alpha = 1.9$, $\beta = 1.0$, $p_u = 0.02$. This concentrates training samples near $t = 1$ where the denoising task is most difficult. Cα and latent channels use independent time schedules: the latent channel uses $\text{Beta}(1.0, 1.5)$ with $p_u = 0.20$, spreading samples more broadly with a bias toward earlier times (near noise).

A.5. Training Configuration

Autoencoder. Optimizer: AdamW with AMSGrad, learning rate 10^{-4}, weight decay 10^{-2}. Loss weights: $\lambda_{\text{struct}} = 1.0$, $\lambda_{\text{seq}} = 1.0$, $\lambda_{\text{KL}} = 10^{-4}$. KL divergence is annealed linearly from 0 to full weight over the first $50{,}000$ training steps. Gradient clipping at global norm 1.0. Precision: bf16-mixed. EMA decay 0.999 over all parameters, swapped in for validation. Batch size 4 per GPU across 6 nodes $\times$ 8 GPUs (48 GPUs total, effective batch size 192). Random seed 42. Trained for 250,000 steps ($\sim$8 epochs, $\sim$15 hours on 6 $\times$ 8 H100 GPUs).

Flow matching. Optimizer: Adam, learning rate 10^{-4}, no weight decay. Loss weights: $\lambda_{\text{CA}} = 1.0$, $\lambda_{\text{latent}} = 1.0$. Self-conditioning with 50% probability (double forward pass). During flow matching training, latent targets are drawn stochastically from the frozen encoder posterior ($\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \cdot \exp(\log \boldsymbol{\sigma})$) rather than using the deterministic mean.Velocity parameterization (v-prediction). CFG categorical mode weights $(0.3, 0.3, 0.3, 0.1)$ for (rama-only, seq-only, coord-only, unconditional). Time sampling: C_α channel $\text{Beta}(1.9, 1.0)$ with 2% uniform mixture; latent channel $\text{Beta}(1.0, 1.5)$ with 20% uniform mixture (independent per-modality schedules following *La-Proteina*). Random SO(3) rotation applied to each structure during training. Gradient clipping at global norm 1.0. Precision: bf16-mixed. EMA decay 0.999 over denoiser parameters only. Same compute configuration as the autoencoder (6 nodes $\times$ 8 GPUs, batch size 4 per GPU). Random seed 42. Trained for 550,000 steps ($\sim$18 epochs, $\sim$38 hours on 6 $\times$ 8 H100 GPUs).

Inference defaults. 200 integration steps. C_α channel: logarithmic time schedule ($p = 2.0$), SDE noise injection $g(t) = 1/t$, noise scaling $T_{\text{CA}} = 0.5$, transition to score-scaled ODE (scale factor 1.5) at $t > 0.98$. Latent channel: power time schedule ($p = 2.0$), SDE noise injection $g(t) = \frac{\pi}{2} \tan\big(\frac{\pi}{2}(1-t)\big)$, noise scaling $T_{\text{lat}} = 0.1$, same ODE transition threshold. Self-conditioning applied at every step (conditional branch predictions fed back). C_α coordinates re-centered to zero center of mass at each step.

B. Dataset Generation

B.1. Sequence Design

Training sequences were generated using a triplet-coverage algorithm that systematically samples local sequence space. For a 39-residue alphabet (19 L-amino acids + 19 D-amino acids + glycine), the total number of ordered three-residue motifs is $39^3 = 59{,}319$. The algorithm guarantees that every one of these triplets appears at least once across the generated sequence set:

1. Initialize the full set of 59,319 triplets as "uncovered."
2. Pop one uncovered triplet as the seed for a new sequence.
3. Extend by appending random residues from the 39-letter alphabet until the sequence reaches target length L.
4. Mark covered: extract all L cyclic triplets $(\text{seq}[i], \text{seq}[(i{+}1) \bmod L], \text{seq}[(i{+}2) \bmod L])$ and remove them from the uncovered set.
5. Repeat steps 2–4 until the uncovered set is empty.

6. Subsample 1,000 sequences uniformly from the covering set.

This procedure was run independently for each ring size ($L = 6, 7, \ldots, 12$), producing 1,000 sequences per length — 7,000 total. Sequences contain mixed L/D residues by construction (no chirality constraints imposed), and no duplicates exist within a given length bin.

B.2. Conformer Generation via Rosetta GenKIC

For each sequence, 1,000 low-energy closed-ring conformers were generated using Rosetta's `simple_cycpep_predict` application (release 4.08) with kinematic closure (GenKIC):

1. Build a linear peptide in extended conformation from the input sequence.
2. Enforce cyclic amide bond constraint between N-terminus of residue 1 and C-terminus of residue L.
3. Sample backbone dihedrals for all residues except three pivot residues, then analytically solve the ring closure equations for the pivots (GenKIC).
4. Reject conformers that failed ring closure or form fewer than 1 backbone hydrogen bond.
5. Energy-minimize accepted conformers with the Rosetta `ref2015` score function.

Key parameters: 10,000 maximum closure attempts per conformer (`-cyclic_peptide:genkic_closure_attempts`), minimum 1 backbone H-bond (`-cyclic_peptide:min_genkic_hbonds`). A wrapper script iteratively calls the application until 1,000 conformers accumulate (up to 10 iterations), then trims to exactly 1,000.

Table 4. Training dataset composition. Two sequences of length 6 failed to produce any conformers within the closure attempt budget. Many smaller cyclic peptides, especially length 6, has fewer than 1000 conformers.

Ring Size	Sequences	Disk Size
L06	998	1.8 GB
L07	1,000	2.1 GB
L08	1,000	2.4 GB
L09	1,000	2.7 GB
L10	1,000	3.0 GB
L11	1,000	3.4 GB
L12	1,000	3.7 GB
Total	**6,998**	**38 GB**

C. D-amino Acid Vocabulary

The model uses a 39-type amino acid vocabulary: 20 canonical amino acids (indices 0-19, one-letter codes A, R, N, …, V) and 19 D-amino acids (indices 20-38, one-letter codes a, r, n, …, v). Each residue is encoded as a 39-dimensional one-hot vector that jointly specifies amino acid identity; chirality is not specified explicitly.

In PDB files, L-amino acids are written as standard ATOM records with canonical three-letter names (ALA, ARG, …); D-amino acids are written as HETATM records with Rosetta naming conventions (DAL, DAR, …). The mapping `D_AA_PARENT_IDX` links each D-amino acid index to its L-parent (20→0, 21→1, …, 38→19, skipping glycine), enabling the decoder to reuse L-parent atom definitions when reconstructing all-atom coordinates for D-residues. Note, several residues have alternate Rosetta names that map to the same index (e.g., DAN → DSG, DCS → DCY, DGU → DGL).

Table 5. Full 39-type amino acid vocabulary. L-amino acids use standard one-letter codes (uppercase); D-amino acids use lowercase. GLY is achiral (no D-form). Rosetta alternate names: DAN=DSG, DCS=DCY, DGU=DGL, DME=MED, DPH=DPN, DSE=DSN.

L-amino acids (+ GLY)				D-amino acids		
#	3-Letter	1-Letter	Full Name	#	3-Letter	1-Letter
0	ALA	A	Alanine	20	DAL	a
1	ARG	R	Arginine	21	DAR	r
2	ASN	N	Asparagine	22	DSG	n
3	ASP	D	Aspartate	23	DAS	d
4	CYS	C	Cysteine	24	DCY	c
5	GLN	Q	Glutamine	25	DGN	q
6	GLU	E	Glutamate	26	DGL	e
7	GLY	G	Glycine			
8	HIS	H	Histidine	27	DHI	h
9	ILE	I	Isoleucine	28	DIL	i
10	LEU	L	Leucine	29	DLE	l
11	LYS	K	Lysine	30	DLY	k
12	MET	M	Methionine	31	MED	m
13	PHE	F	Phenylalanine	32	DPN	f
14	PRO	P	Proline	33	DPR	p
15	SER	S	Serine	34	DSN	s
16	THR	T	Threonine	35	DTH	t
17	TRP	W	Tryptophan	36	DTR	w
18	TYR	Y	Tyrosine	37	DTY	y
19	VAL	V	Valine	38	DVA	v

D. Evaluation Metrics

We evaluate ensemble fidelity using seven complementary metrics that assess distributional similarity, conformational diversity, and sequence prediction accuracy. For each test sequence, the model generates 1,000 conformers; ground truth (GT) consists of up to 1,000 Rosetta-generated conformers.

D.1. Ramachandran cosine similarity (RamaCos ↑).

For each residue position j, we build a 2D (ϕ_j, ψ_j) histogram with 80×80 bins over $[-180°, 180°)$ ($4.5°$ resolution) from the generated ensemble (gen) and the ground-truth ensemble (gt). Each histogram is L1-normalized to a probability vector $\mathbf{p}_j^{\text{gen}}, \mathbf{p}_j^{\text{gt}} \in \mathbb{R}^{6400}$. Per-residue cosine similarity is:

$$\text{RamaCos}_j = \frac{\mathbf{p}_j^{\text{gen}} \cdot \mathbf{p}_j^{\text{gt}}}{\|\mathbf{p}_j^{\text{gen}}\|_2 \, \|\mathbf{p}_j^{\text{gt}}\|_2} \tag{6}$$

The reported value is the mean over all n residue positions: $\text{RamaCos} = \frac{1}{n}\sum_{j=1}^{n} \text{RamaCos}_j$. Dihedral angles are extracted with cyclic wrapping: ϕ_j uses atoms $(C_{(j-1) \bmod n},\, N_j,\, C_j^{\alpha},\, C_j)$ and ψ_j uses $(N_j,\, C_j^{\alpha},\, C_j,\, N_{(j+1) \bmod n})$. Values range from 0 (orthogonal distributions) to 1 (identical distributions).

D.2. Cluster code cosine similarity (ClusterCos ↑).

Each residue's (ϕ, ψ) is assigned to the nearest of 12 data-driven conformational clusters (6 L-form + 6 D-form centers, determined by k-means in $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ space on a 50k-conformer subsample from the training set, see Appendix E). Assignment uses Euclidean distance in the 4D $\cos$ / $\sin$ embedding. Each conformer is then represented as a string of n cluster labels (one Greek letter per residue), yielding a full-peptide conformational *code*.

The frequency distribution over all unique codes is computed for generated (gen) and Rosetta (gt) ensembles independently, yielding probability vectors $\mathbf{q}^{\text{gen}}, \mathbf{q}^{\text{gt}}$ over the union of observed codes. ClusterCos is the cosine similarity between these vectors:

$$\text{ClusterCos} = \frac{\mathbf{q}^{\text{gen}} \cdot \mathbf{q}^{\text{gt}}}{\|\mathbf{q}^{\text{gen}}\|_2 \, \|\mathbf{q}^{\text{gt}}\|_2} \tag{7}$$

This metric captures whether the model reproduces the correct conformational substates *and* their relative populations.

Because the number of possible codes grows as 12^n, ClusterCos decreases systematically with peptide length even for well performing models (longer sequences have sparser code distributions; possibly, even more accessible conformations than have been generated).

D.3. PCA sliced Wasserstein distance (PCA-SW ↓).

Each conformer's (ϕ, ψ) angles are transformed to a $4n$-dimensional feature vector via $[\cos\phi_1, \ldots, \cos\phi_n, \sin\phi_1, \ldots, \sin\phi_n, \cos\psi_1, \ldots, \cos\psi_n, \sin\psi_1, \ldots, \sin\psi_n]$ (the $\cos / \sin$ embedding avoids the $\pm 180^\circ$ discontinuity). PCA (2 components) is fit on the combined generated + Rosetta (GT) feature matrix, projecting both ensembles into a shared 2D subspace.

The sliced Wasserstein distance is then computed between the two 2D point clouds:

$$\text{PCA-SW} = \frac{1}{K} \sum_{k=1}^{K} W_1\big(\pi_{\theta_k} P^{\text{gen}},\ \pi_{\theta_k} P^{\text{gt}}\big) \tag{8}$$

where π_{θ_k} denotes projection onto a random 1D direction θ_k, W_1 is the 1-Wasserstein distance on the projected 1D distributions, and $K = 50$ random directions are sampled (seed fixed for reproducibility). Lower values indicate better distributional overlap in the principal conformational subspace.

D.4. PCA cosine similarity (PCA-CosSim ↑).

Using the same PCA projection as PCA-SW, the 2D point clouds are binned into 80×80 2D histograms over the data range (with 5% padding). The resulting count vectors are L1-normalized to probability distributions, and cosine similarity is computed between them. This complements PCA-SW by measuring directional alignment of the conformational distributions in PC space rather than earth-mover distance.

D.5. Sequence recovery (SeqRec ↑).

The autoencoder decoder predicts an amino acid sequence (from the 39-type vocabulary including D-amino acids) for each generated conformer. Since each test case corresponds to a single peptide sequence s_{GT}, SeqRec is simply the fraction of generated conformers whose decoded sequence matches it exactly:

$$\text{SeqRec} = \frac{1}{N_{\text{gen}}} \sum_{i=1}^{N_{\text{gen}}} \mathbf{1}\big[\hat{s}_i = s_{\text{GT}}\big] \tag{9}$$

where $\hat{s}_i$ is the decoded sequence of generated conformer i. This requires all n positions to match exactly, including L/D chirality. A conformer that decodes to the wrong sequence is not a valid structure for the target peptide, so SeqRec measures physical consistency of the generated ensemble.

D.6. Per-residue recovery (PRR ↑).

The plurality (most frequent) decoded sequence among the 1,000 generated conformers is compared position-by-position against the ground-truth sequence:

$$\text{PRR} = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}\big[\hat{s}^{\star}_j = s_{\text{GT},j}\big] \tag{10}$$

where $\hat{s}^{\star} = \arg\max_s \text{count}(s \mid \text{Gen})$ is the top-1 generated sequence and s_{GT} is the ground-truth sequence. Unlike SeqRec (which requires an exact full-sequence match), PRR measures local per-position accuracy of the single most populated prediction.

D.7. Ensemble entropy (Entropy).

Shannon entropy over the cluster code frequency distribution, restricted to the top-100 most frequent codes and renormalized:

$$H = -\sum_{k=1}^{\min(|\mathcal{C}|,100)} \tilde{p}_k \ln \tilde{p}_k, \qquad \tilde{p}_k = \frac{p_k}{\sum_{k'=1}^{100} p_{k'}} \tag{11}$$

where $\{p_k\}$ are the frequencies of full-peptide cluster codes sorted in decreasing order. We compute entropy separately for generated (H_{gen}) and ground-truth (H_{gt}) ensembles. The top-100 truncation ensures the metric is dominated by meaningfully populated states rather than noise in the long tail. Higher values indicate greater conformational diversity. For a uniform distribution over m codes, $H = \ln m$; the maximum for 100 codes is $\ln(100) \approx 4.61$.

D.8. Sequence uniqueness (SeqUnique).

The number of distinct amino acid sequences decoded from the 1,000 generated conformers:

$$\text{SeqUnique} = |\{\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_{1000}\}| \tag{12}$$

Low uniqueness (close to 1) indicates the model collapses to a single sequence prediction regardless of conformer geometry—expected and desirable under sequence conditioning. High uniqueness (close to 1,000) indicates every conformer is decoded to a different sequence—expected under coord-only or unconditional generation where no sequence signal is provided. This diagnostic distinguishes between modes where the model has learned to map structure to a specific sequence versus modes in which it has not.

D.9. Ring closure.

For cyclic peptides, the backbone must form a closed ring via a peptide bond between C of the last residue and N of the first residue. We measure the Euclidean distance $d(\text{C}_{\text{last}}, \text{N}_{\text{first}})$ in Angstroms for each generated conformer. An ideal peptide bond has $d \approx 1.33$ Å; conformers with $d > 1.5$ Å are classified as "open" (failed ring closure). The fraction of open conformers indicates whether the model has learned the geometric constraint of ring cyclization without explicit enforcement during training or inference.

E. Conformational Cluster Codes

To compactly represent per-residue backbone conformations, we define a 12-state cluster code system derived from data-driven Voronoi tessellation of the Ramachandran plane.

E.1. Tessellation Procedure

We subsample 50,000 conformers (totaling 451,659 ϕ, ψ dihedral angle pairs) uniformly across all peptide lengths and extract per-residue (ϕ, ψ) dihedral angles. Each angle pair is embedded in a 4D $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ representation that respects the angular periodicity. We then run k-means clustering (k=6) separately on L-amino acid residues (negative-ϕ region) and D-amino acid residues (positive-ϕ region), yielding 6 L-form and 6 D-form cluster centers for a total of 12 states.

The 12 clusters partition the Ramachandran plane into Voronoi regions in the 4D $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ metric space (Figure 5). Each residue in a conformer is assigned to the nearest cluster center by Euclidean distance in this space, producing a per-residue single-character code. L-form clusters are denoted by uppercase letters (A, B, Γ, Δ, E, Z) and D-form clusters by lowercase (α, β, γ, δ, ϵ, ζ). A conformer of an n-residue peptide is thus encoded as an n-character string.

E.2. Cluster Centers

Table 6 lists the 12 cluster centers in (ϕ, ψ) coordinates. Each D-form cluster is the approximate mirror image of its L-form counterpart, with $\phi_{\text{D}} \approx -\phi_{\text{L}}$ and $\psi_{\text{D}} \approx -\psi_{\text{L}}$.

L-form clusters				D-form clusters			
#	Name	ϕ	ψ	#	Name	ϕ	ψ
1	A	−68.0	−30.5	7	α	68.2	30.6
2	B	−103.5	7.4	8	β	103.2	−7.1
3	Γ	−65.6	133.2	9	γ	64.9	−133.1
4	Δ	−139.3	154.2	10	δ	140.5	−155.8
5	E	−92.1	162.9	11	ϵ	92.7	−160.0
6	Z	−91.7	76.7	12	ζ	92.1	−77.1

Table 6. Cluster center coordinates (degrees) from k-means in $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ space.

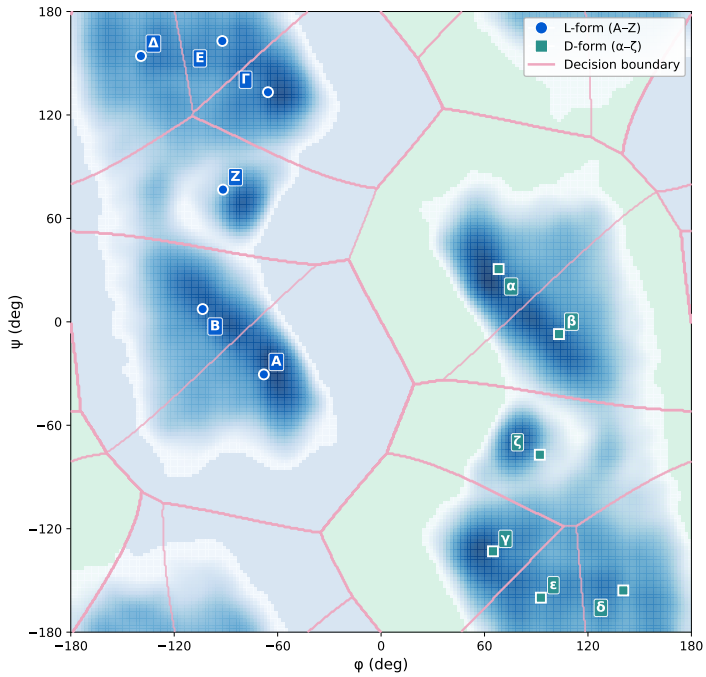

Figure 5. Decision boundaries of the 12-state conformational cluster system, computed in the $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ metric space and projected onto the (ϕ, ψ) plane. Boundaries are curved because equidistant surfaces in the periodic cos/sin embedding are non-linear in degree space. Blue circles: L-form cluster centers ($\phi < 0$); green squares: D-form centers ($\phi > 0$). Background: smoothed density from 451,659 residue observations.

E.3. Usage in Evaluation

The cluster code cosine similarity (ClusterCos) metric compares the frequency distributions of full-peptide conformational codes between generated and ground-truth ensembles. For a peptide of length n, each conformer maps to an n-character string; cosine similarity between the frequency vectors of these strings measures whether the model reproduces both the correct conformational substates and their relative populations.

Length dependence. ClusterCos shows strong length dependence: hexapeptides (6-mers) achieve ClusterCos > 0.7 (a 6-character code has $12^6 \approx 3 \times 10^6$ possible states, but most peptides populate only 5–20 dominant codes), while 12-mers drop to near zero because $12^{12} \approx 9 \times 10^{12}$ possible states (many of which are thermally accessible) makes exact code matches vanishingly rare even between correctly generated ensembles. For longer peptides, per-residue metrics (RamaCos) are more informative than full-peptide code metrics.

F. Ramachandran Plot Examples

F.1. Sample Cyclic Peptide Ramachandran Plots

Per-residue ϕ,ψ-backbone distributions from the ground-truth (Rosetta) conformer ensembles for one representative cyclic peptide at each length from 6-mers to 12-mer (L06-L12). Each panel shows the (ϕ, ψ) density over 1,000 conformers for a single residue position; column headers indicate the amino acid type (uppercase = L, lowercase = D).

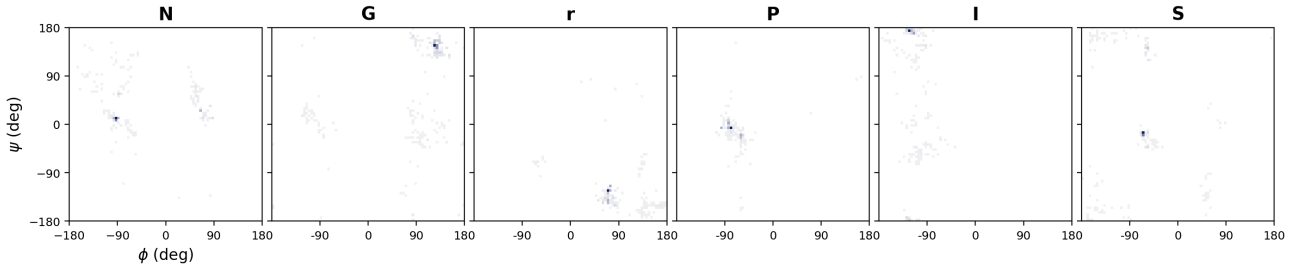

Figure 6. Per-residue Ramachandran distributions for a representative hexapeptide (L06).

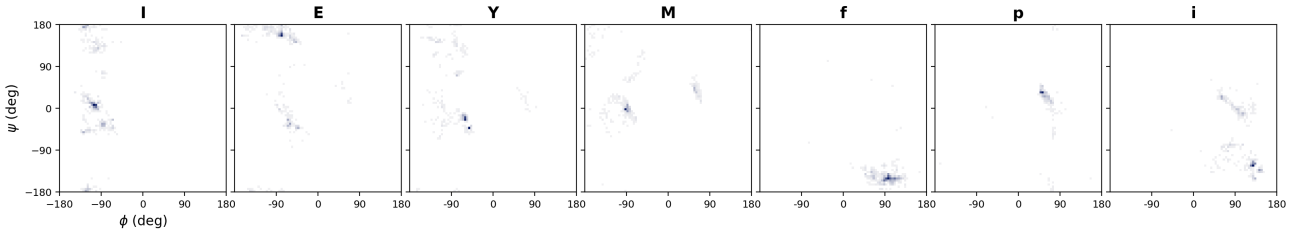

Figure 7. Per-residue Ramachandran distributions for a representative heptapeptide (L07).

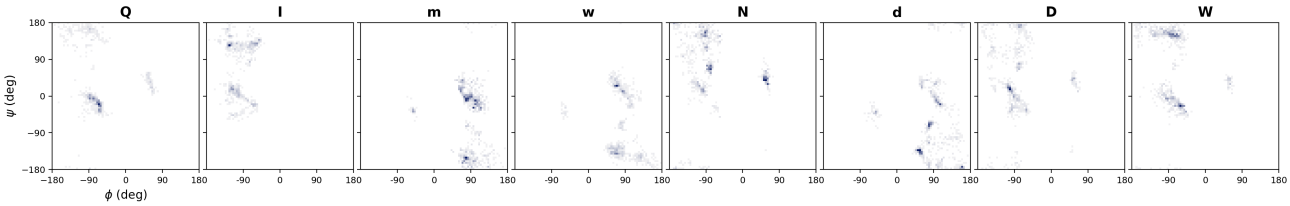

Figure 8. Per-residue Ramachandran distributions for a representative octapeptide (L08).

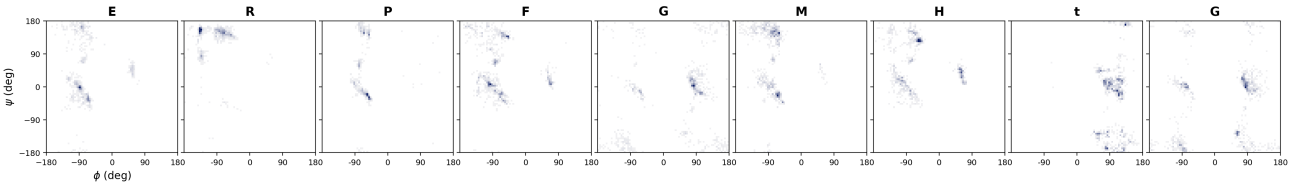

Figure 9. Per-residue Ramachandran distributions for a representative nonapeptide (L09).

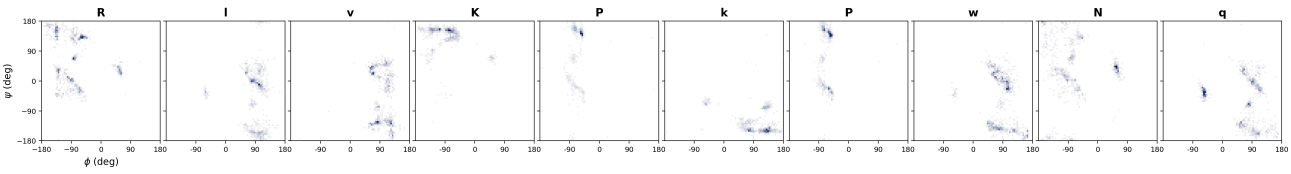

Figure 10. Per-residue Ramachandran distributions for a representative decapeptide (L10).

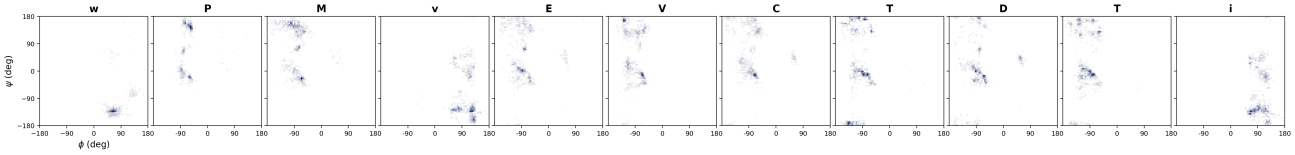

Figure 11. Per-residue Ramachandran distributions for a representative undecapeptide (L11).

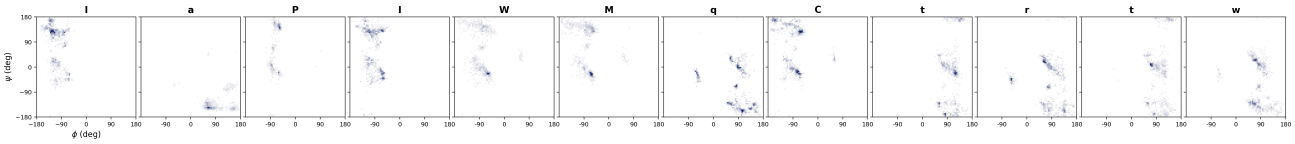

Figure 12. Per-residue Ramachandran distributions for a representative dodecapeptide (L12).

F.2. Homopolymer Ramachandran Distributions

As a baseline, we generate 2,500 conformers of cyclic homo-L-alanine (Cyclo-A) at each ring size using the same Rosetta GenKIC protocol. Because all residues are identical, the per-residue Ramachandran distributions reveal how ring geometry alone constrains backbone conformations — independent of sequence-dependent sidechain effects.

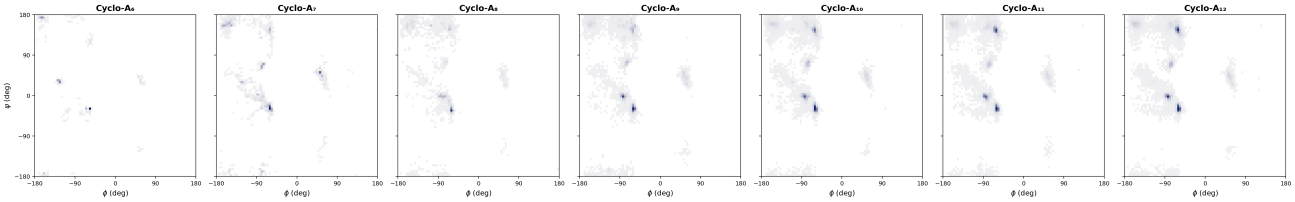

Figure 13. Per-residue Ramachandran distributions for cyclic homo-L-alanine (Cyclo-A_6 through Cyclo-A_{12}), aggregated over all residue positions and 2,500 conformers per ring size. Shorter rings (L06–L08) are restricted to a small number of tight clusters in the α_R and β regions; longer rings (L10–L12) progressively populate additional conformational states as the ring geometry becomes less constraining.

F.3. Per-Residue-Type Ramachandran Distributions

Figures 14 and 15 show the Ramachandran distributions aggregated by amino acid type across the entire training set (all ring sizes, all sequence positions). Each panel aggregates (ϕ, ψ) observations for one residue type over all occurrences in the dataset. L-amino acids predominantly occupy the α_R ($\phi \approx -60°$, $\psi \approx -45°$) and β ($\phi \approx -120°$, $\psi \approx 130°$) basins, while glycine — with no sidechain — populates both left- and right-handed regions symmetrically. D-amino acids show the mirror-image pattern, with density concentrated in the positive-ϕ half of the Ramachandran plane.

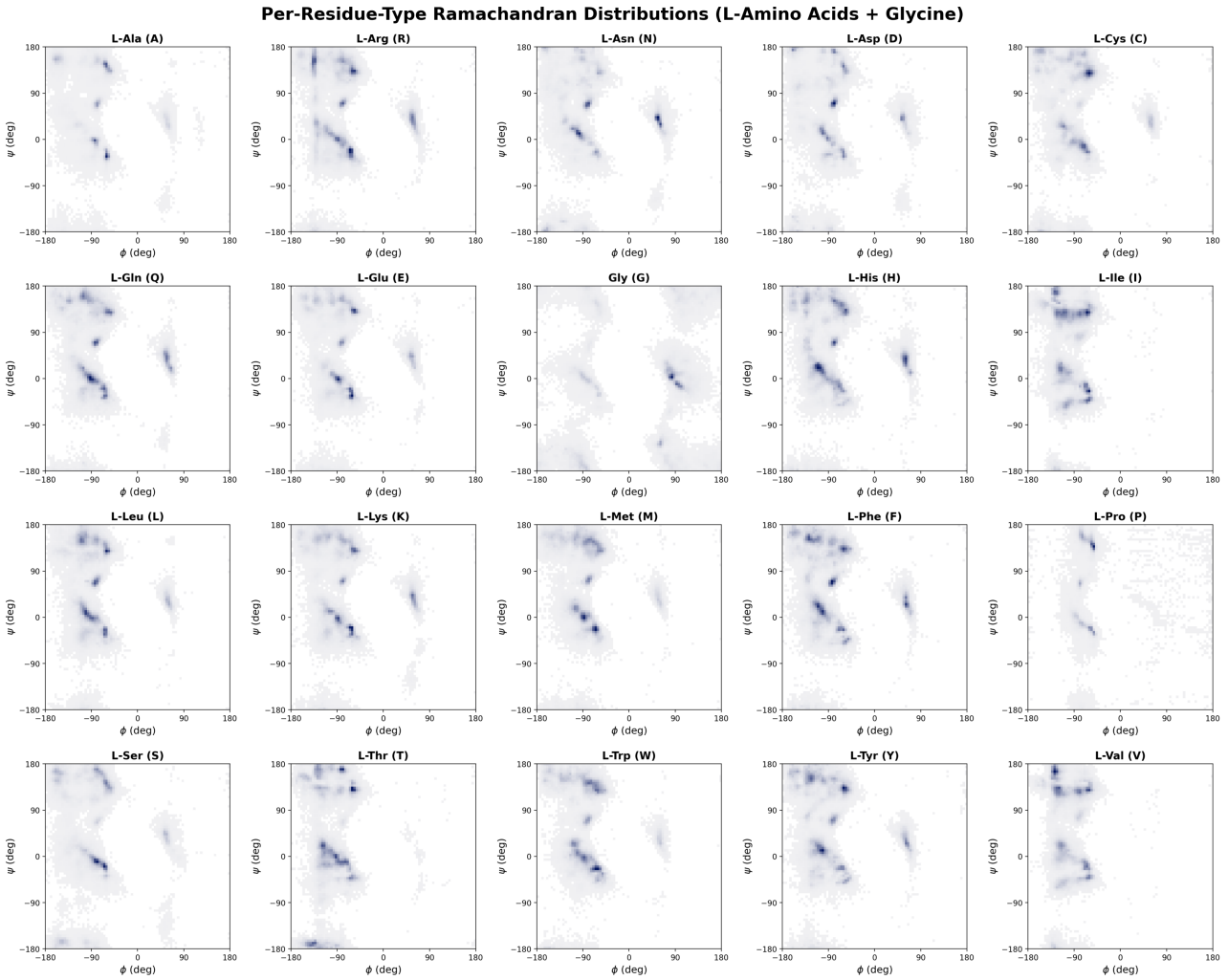

Figure 14. Per-residue-type Ramachandran distributions for L-amino acids and glycine (20 types), aggregated across all training set conformers. Each panel shows the (ϕ, ψ) density for all occurrences of that residue type in the dataset.

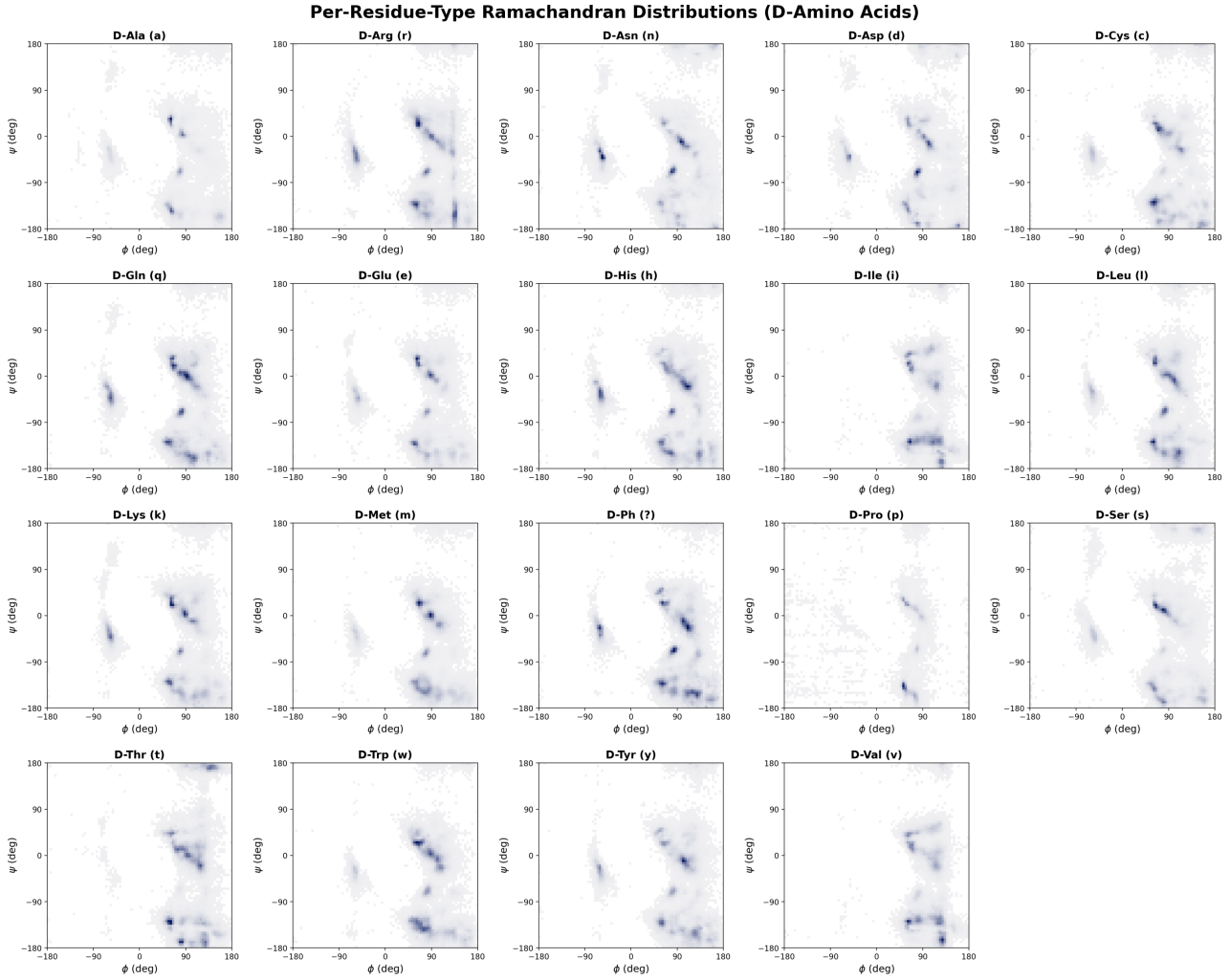

Figure 15. Per-residue-type Ramachandran distributions for D-amino acids (19 types). Distributions are approximate mirror images of the corresponding L-amino acids, with density concentrated in the positive-ϕ region.

G. Generation Speed

We benchmark generation throughput on a single H100 GPU using the full two-stage pipeline (flow matching denoiser + autoencoder decoder) with default inference settings: 200 integration steps, logarithmic time schedule, SDE sampling mode ("sc"), T_{CA}=0.5, T_{lat}=0.1. Batch sizes of 10, 100, and 1,000 conformers are tested across peptide lengths 6, 9, and 12. Each configuration is timed over 5 repeats after 3 warmup runs to ensure stable GPU kernel caching. Timing includes both the denoiser integration (200 forward passes) and the autoencoder decode step; CUDA synchronization is enforced before each measurement. At batch size 1,000, the model generates hexapeptide (L06) conformers at 82 samples/sec (0.012 sec/sample) and 12-mer (L12) conformers at 40 samples/sec (0.025 sec/sample). At small batch size (N=10), per-sample latency is dominated by fixed overhead ($\sim$0.22 sec/sample regardless of length), while at N=1,000 amortized throughput scales efficiently. Compared to Rosetta GenKIC, which requires $\sim$1–5 sec per hexapeptide conformer and $\sim$10–30 sec per 12-mer on a single CPU core, CPGEN achieves $\sim$80$\times$ and $\sim$400$\times$ speedups respectively at batch size 1,000 on a single GPU.

Table 7. Generation speed benchmark. Full two-stage pipeline (denoiser + decoder), 200 integration steps, single GPU, 5 repeats after 3 warmup runs.

Batch	Length	Time (s)	Samples/sec	Sec/sample
10	6	2.18 ± 0.04	4.6	0.218
100	6	2.21 ± 0.00	45.2	0.022
1000	6	12.18 ± 0.04	82.1	0.012
10	9	2.15 ± 0.01	4.6	0.215
100	9	2.86 ± 0.01	34.9	0.029
10	12	2.16 ± 0.00	4.6	0.216
100	12	3.27 ± 0.00	30.6	0.033
1000	12	24.72 ± 0.03	40.5	0.025

Table 8. Comparison to Rosetta GenKIC at batch size N=1,000 (single GPU vs. single CPU core).

Method	L06 (sec/conformer)	L12 (sec/conformer)
Rosetta GenKIC (1 CPU core)	$\sim$1–5	$\sim$10–30
CPGEN (1 GPU, N=1000)	0.012	0.025
Speedup	$\sim$80–400$\times$	$\sim$400–1200$\times$

Three observations emerge from the benchmark. First, fixed overhead dominates at small batch sizes: at N=10, all lengths take $\sim$2.2s total ($\sim$0.22 sec/sample), indicating that the 200 denoiser forward passes are the bottleneck regardless of sequence length. Second, batch scaling is efficient: going from N=10 to N=1,000, per-sample cost drops $\sim$18$\times$ for L06 (0.218 $\rightarrow$ 0.012) and $\sim$9$\times$ for L12 (0.216 $\rightarrow$ 0.025). Third, length scaling is modest: at N=1,000, L12 is only $\sim$2$\times$ slower than L06 (24.7s vs 12.2s), because the transformer's quadratic attention cost is negligible at these short sequence lengths.

G.1. Rosetta GenKIC Computational Cost

Table 9 reports the wall-clock time required by Rosetta `simple_cycpep_predict` (GenKIC) to generate 1,000 conformers per sequence on a single CPU core (Intel Xeon Ice Lake, c7i instances). Timing statistics are computed from 200 representative sequences per ring size.

Table 9. Rosetta GenKIC wall-clock time per sequence (1,000 conformers, single CPU core). High variance at L06 reflects sequence-dependent closure difficulty — short rings with bulky residues (Trp, Pro) or all-L/all-D stretches require many more closure attempts.

Ring Size	Mean (h)	Median (h)	Std (h)	P10 (h)	P90 (h)	Total CPU-h
L06	2.3	1.7	2.1	0.5	4.8	2,349
L07	0.9	0.7	0.6	0.5	1.4	916
L08	0.8	0.7	0.4	0.5	1.1	805
L09	1.0	0.9	0.5	0.7	1.4	997
L10	1.3	1.2	0.6	0.8	1.9	1,278
L11	1.4	1.3	0.7	1.0	2.0	1,444
L12	1.6	1.5	0.6	1.1	2.1	1,562
All						**9,351**

The total dataset generation required $\sim$9,400 CPU-hours ($\sim$390 CPU-days), executed over 3.5 days of wall-clock time by running $\sim$7,000 single-core SLURM jobs in parallel on an AWS ParallelCluster. L06 is paradoxically the most expensive ring size per sequence: fewer backbone degrees of freedom make it geometrically harder to find valid GenKIC solutions, wasting many closure attempts before producing an accepted conformer. L08 represents the computational sweet spot (fewest wasted attempts); cost then increases monotonically with length due to the larger number of atoms requiring energy minimization after closure.

By comparison, CPGEN generates 1,000 L06 conformers in 12.2 seconds and 1,000 L12 conformers in 24.7 seconds on

a single GPU — replacing $\sim$2.3 hours and $\sim$1.6 hours of CPU time respectively. Across the full dataset, the $\sim$7 million conformers that required 9,400 CPU-hours with Rosetta could be regenerated in $\sim$24 GPU-hours with CPGEN ($\sim$390$\times$ overall speedup).

H. Test Set Results

Test set evaluation on 50 held-out sequences per length (350 total), 1,000 conformers each. Two parameter configurations: **Config A** (w=0.75, T_{CA}=0.5, T_{lat}=0.1, optimized) and **Config B** (w=1.0, T_{CA}=0.1, T_{lat}=0.1, original defaults).

Table 10. Test set results — seq_only, Config A (w=0.75, T_{CA}=0.5, T_{lat}=0.1). 50 sequences/length, 1,000 conformers each.

Metric	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	0.524	0.403	0.453	0.528	0.566	0.576	0.576	0.570
ClusterCos ↑	0.287	0.725	0.504	0.424	0.252	0.085	0.017	0.003
PCA-CosSim ↑	0.365	0.403	0.417	0.472	0.401	0.323	0.273	0.268
PCA-SW ↓	0.268	0.441	0.369	0.265	0.226	0.208	0.191	0.178
SeqRec ↑	0.998	0.999	0.992	0.999	0.998	0.998	0.998	0.998
PRR ↑	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 11. Test set results — rama_only, Config A (w=0.75, T_{CA}=0.5, T_{lat}=0.1). 50 sequences/length, 1,000 conformers each.

Metric	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	0.526	0.392	0.451	0.537	0.568	0.580	0.580	0.572
ClusterCos ↑	0.286	0.699	0.507	0.426	0.253	0.093	0.018	0.003
PCA-CosSim ↑	0.363	0.367	0.429	0.483	0.399	0.326	0.275	0.264
PCA-SW ↓	0.287	0.442	0.396	0.319	0.243	0.219	0.204	0.187
SeqRec ↑	0.604	0.059	0.363	0.557	0.735	0.736	0.863	0.913
PRR ↑	0.899	0.600	0.823	0.930	0.976	0.970	0.993	1.000

Table 12. Test set results — coords_only, Config A (w=0.75, T_{CA}=0.5, T_{lat}=0.1). 50 sequences/length, 1,000 conformers each. Conditioned on 100 randomly selected ground-truth C_α conformers per sequence.

Metric	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	0.544	0.463	0.507	0.550	0.569	0.578	0.573	0.568
ClusterCos ↑	0.372	0.833	0.647	0.484	0.272	0.171	0.107	0.092
PCA-CosSim ↑	0.382	0.512	0.517	0.502	0.382	0.296	0.237	0.230
PCA-SW ↓	0.149	0.163	0.138	0.144	0.150	0.139	0.147	0.162
SeqRec ↑	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRR ↑	0.118	0.133	0.131	0.118	0.129	0.100	0.087	0.127

Table 13. Config A vs Config B — seq_only. Bold indicates the better configuration.

Metric	Config	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	A	**0.524**	**0.403**	**0.453**	**0.528**	**0.566**	**0.576**	**0.576**	**0.570**
	B	0.477	0.360	0.389	0.462	0.504	0.532	0.542	0.548
ClusterCos ↑	A	**0.287**	**0.725**	**0.504**	**0.424**	**0.252**	0.085	0.017	0.003
	B	0.250	0.648	0.422	0.342	0.212	**0.096**	**0.027**	**0.006**
PCA-CosSim ↑	A	**0.365**	**0.403**	**0.417**	**0.472**	**0.401**	**0.323**	**0.273**	**0.268**
	B	0.276	0.341	0.340	0.381	0.285	0.233	0.175	0.177
PCA-SW ↓	A	**0.268**	**0.441**	**0.369**	**0.265**	**0.226**	**0.208**	**0.191**	**0.178**
	B	0.649	0.741	0.731	0.711	0.620	0.603	0.592	0.548
SeqRec ↑	A	0.998	0.999	0.992	0.999	0.998	0.998	0.998	0.998
	B	**0.999**	**1.000**	**0.995**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Table 14. Config A vs Config B — rama_only. Bold indicates the better configuration.

Metric	Config	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	A	**0.526**	**0.392**	**0.451**	**0.537**	**0.568**	**0.580**	**0.580**	**0.572**
	B	0.473	0.345	0.385	0.459	0.497	0.533	0.540	0.552
ClusterCos ↑	A	**0.286**	**0.699**	**0.507**	**0.426**	**0.253**	0.093	0.018	0.003
	B	0.251	0.622	0.428	0.358	0.214	**0.103**	**0.027**	**0.007**
PCA-CosSim ↑	A	**0.363**	**0.367**	**0.429**	**0.483**	**0.399**	**0.326**	**0.275**	**0.264**
	B	0.272	0.322	0.346	0.372	0.286	0.229	0.175	0.176
PCA-SW ↓	A	**0.287**	**0.442**	**0.396**	**0.319**	**0.243**	**0.219**	**0.204**	**0.187**
	B	0.709	0.784	0.812	0.818	0.704	0.659	0.611	0.577
SeqRec ↑	A	0.604	0.059	0.363	0.557	0.735	0.736	0.863	0.913
	B	**0.609**	**0.069**	**0.375**	**0.558**	**0.742**	**0.737**	**0.869**	**0.916**
PRR ↑	A	0.899	0.600	**0.823**	**0.930**	0.976	0.970	**0.993**	**1.000**
	B	**0.900**	**0.603**	0.823	0.930	**0.978**	**0.976**	0.991	1.000

Table 15. Config A vs Config B — coords_only. Bold indicates the better configuration.

Metric	Config	ALL	L06	L07	L08	L09	L10	L11	L12
RamaCos ↑	A	**0.544**	**0.463**	**0.507**	**0.550**	**0.569**	**0.578**	**0.573**	**0.568**
	B	0.533	0.459	0.498	0.540	0.557	0.565	0.559	0.556
ClusterCos ↑	A	**0.372**	0.833	**0.647**	**0.484**	**0.272**	**0.171**	**0.107**	**0.092**
	B	0.365	**0.838**	0.643	0.462	0.260	0.163	0.102	0.088
PCA-CosSim ↑	A	**0.382**	**0.512**	**0.517**	**0.502**	**0.382**	**0.296**	0.237	**0.230**
	B	0.371	0.507	0.502	0.484	0.363	0.283	**0.234**	0.226
PCA-SW ↓	A	0.149	**0.163**	0.138	0.144	0.150	0.139	0.147	0.162
	B	**0.147**	0.164	**0.137**	**0.143**	**0.148**	**0.136**	**0.144**	**0.159**
PRR ↑	A	0.118	0.133	0.131	0.118	0.129	0.100	0.087	0.127
	B	**0.138**	**0.163**	**0.140**	**0.130**	**0.138**	**0.124**	**0.131**	**0.137**

Table 16. Sequence recovery from rama-only conditioning (Config A). Exact match = fraction of 1,000 conformers with correct full sequence; per-residue = fraction of residue positions correct in the top-1 predicted sequence.

Metric	ALL	L06	L07	L08	L09	L10	L11	L12
Exact match (%)	60.4	5.9	36.3	55.7	73.5	73.6	86.3	91.3
Per-residue (%)	89.9	60.0	82.3	93.0	97.6	97.0	99.3	100.0

H.1. Test Set Ramachandran Plots

Per-residue Ramachandran plots for the top-2 highest RamaCos sequences at each of three lengths (L06, L12) from the test set, using optimized parameters (Config A: w=0.75, T_{CA}=0.5, T_{lat}=0.1). Top row: ground truth (GT, from Rosetta). Bottom row: generated ensemble (1,000 conformers). All plots use 80×80 bins (4.5° resolution).

H.1.1. HEXAPEPTIDES (6-MER) USING SEQ-ONLY SAMPLING MODE

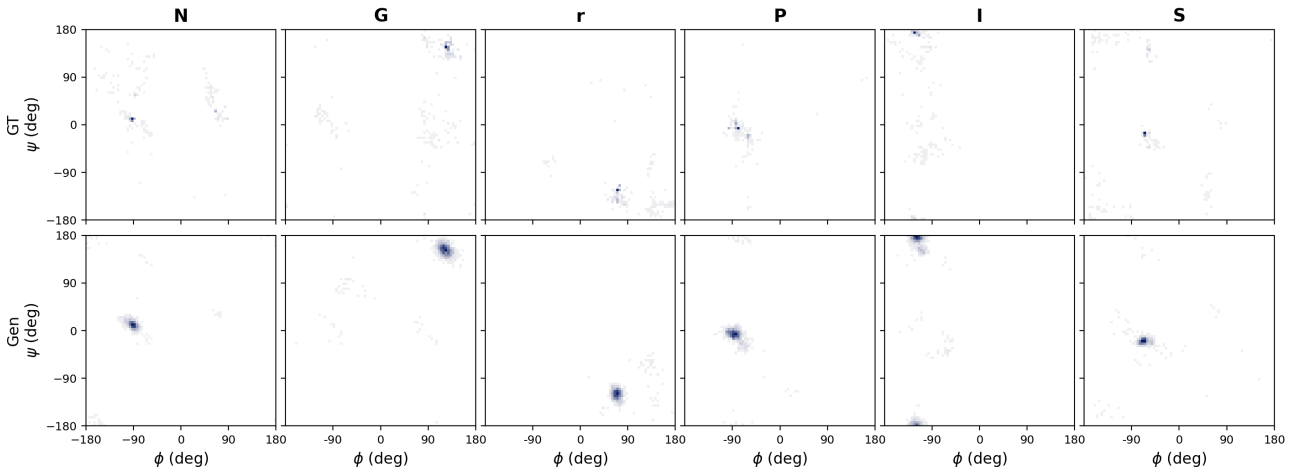

Figure 16. Sequence: NGrPIS, RamaCosSim = 0.488

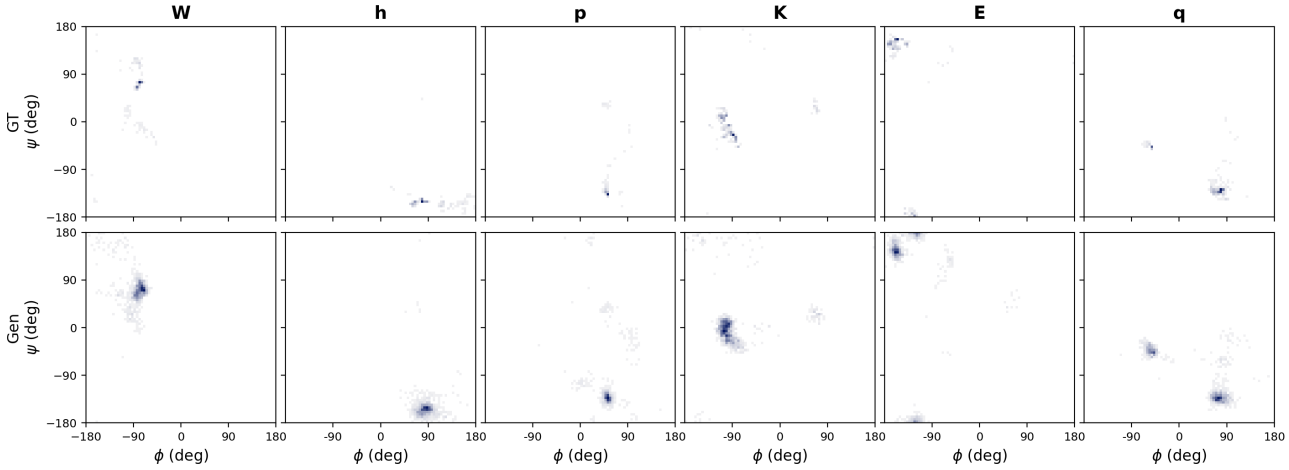

Figure 17. Sequence: WhpKEq, RamaCosSim = 0.486

H.1.2. 12-MER USING SEQ-ONLY SAMPLING MODE

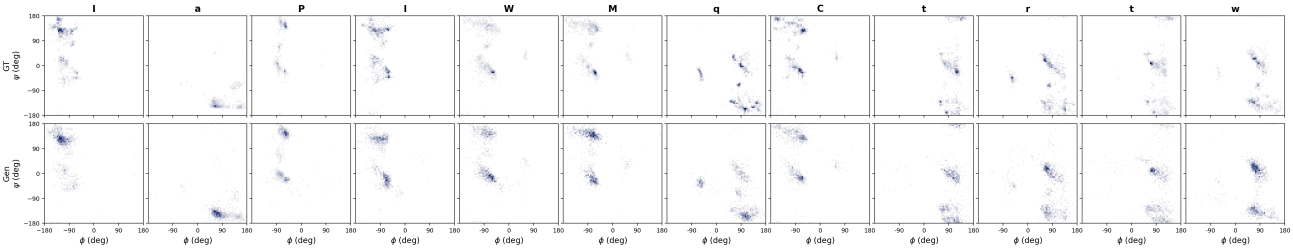

Figure 18. Sequence: IaPIWMqCtrtw, RamaCosSim = 0.616

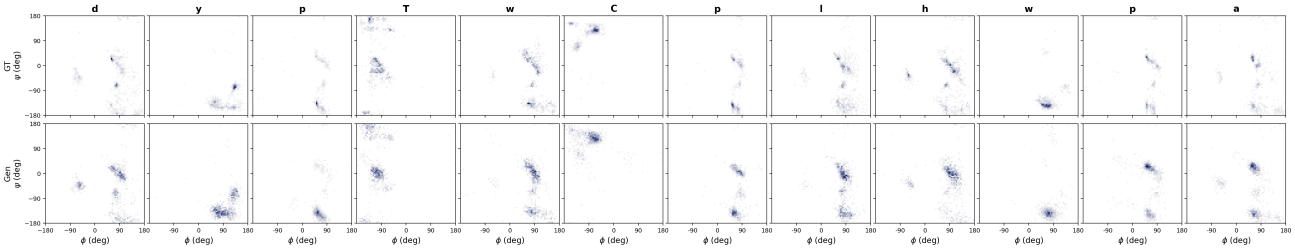

Figure 19. Sequence: dypTwCplhwpa, RamaCosSim = 0.608

H.2. Test Set PCA Comparison Plots

PCA projections of $(\cos\phi, \sin\phi, \cos\psi, \sin\psi)$ features for generated (blue) vs. ground-truth (pink) ensembles, for the same top-scoring test sequences. Overlap between the two point clouds indicates distributional agreement in the global conformational landscape beyond individual residue Ramachandran similarity.

H.2.1. Hexapeptides (L06)

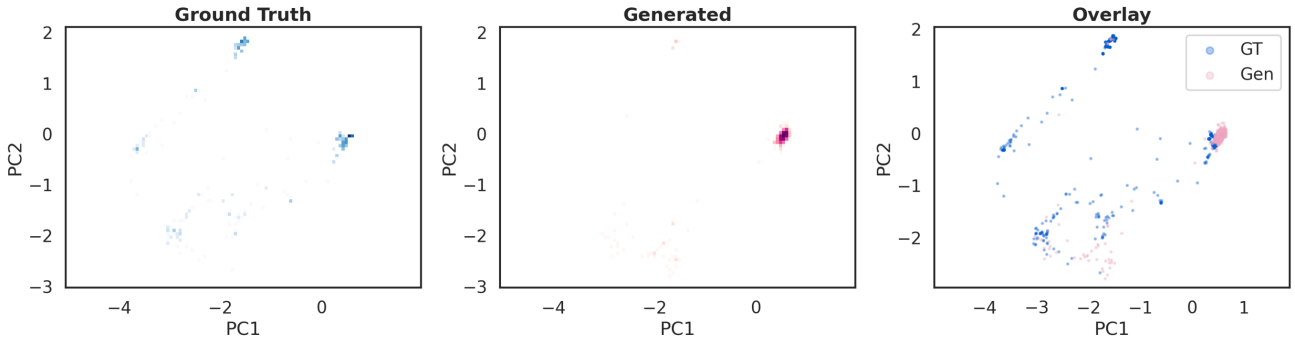

Figure 20. Sequence: NGrPIS, PCA-CosSim = 0.730, PCA-SW = 0.627

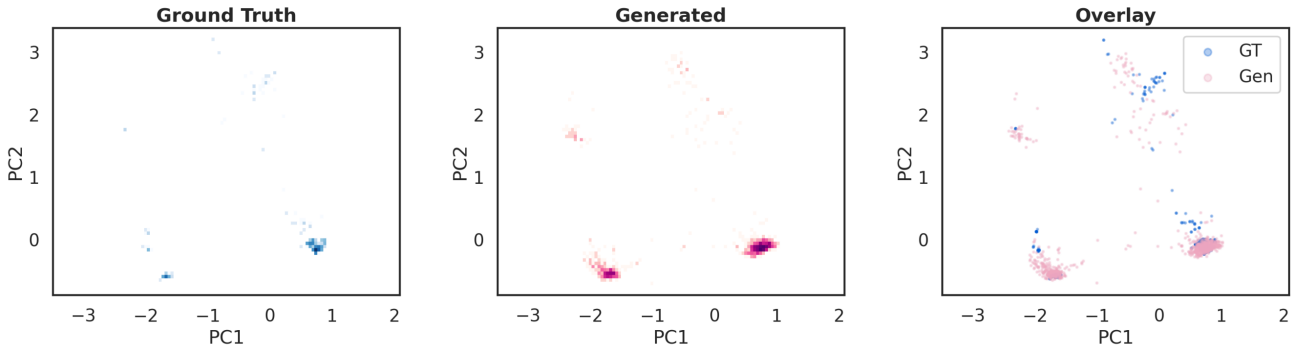

Figure 21. Sequence: WhpKEq, PCA-CosSim = 0.658, PCA-SW = 0.324

H.2.2. 9-MERS (L09)

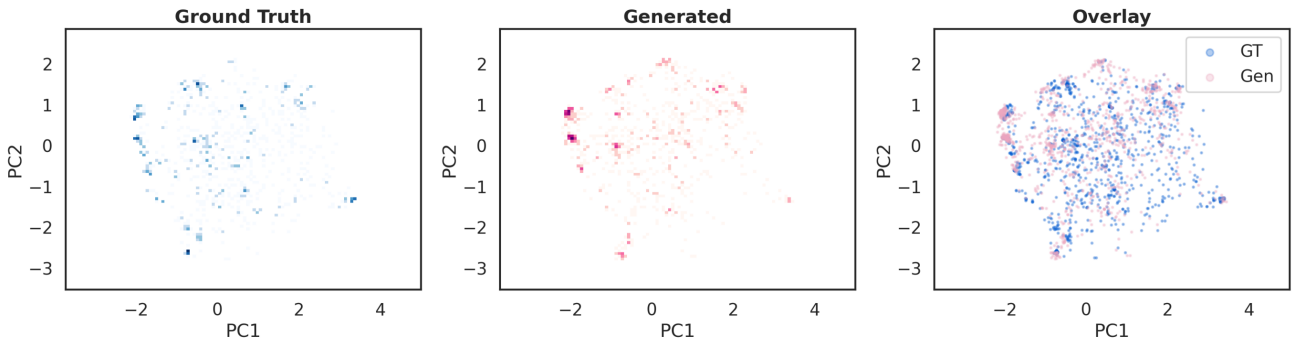

Figure 22. Sequence: ERPFGMHtG, PCA-CosSim = 0.351, PCA-SW = 0.314

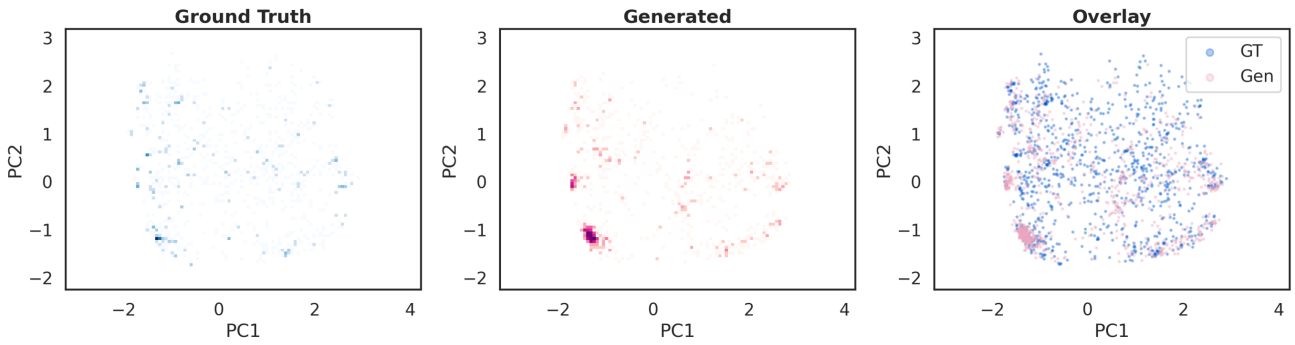

Figure 23. Sequence: NNlAQRqpY, PCA-CosSim = 0.410, PCA-SW = 0.355

H.2.3. 12-MERS (L12)

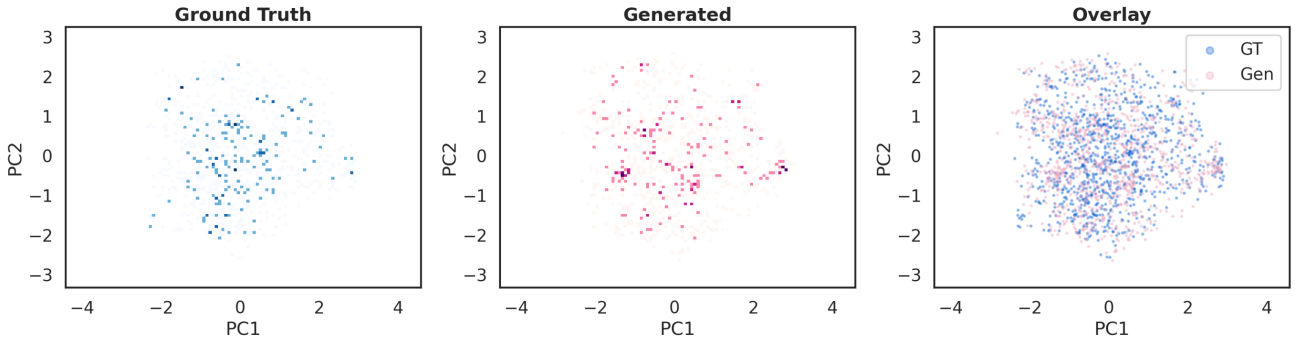

Figure 24. Sequence: IaPIWMqCtrtw, PCA-CosSim = 0.260, PCA-SW = 0.088

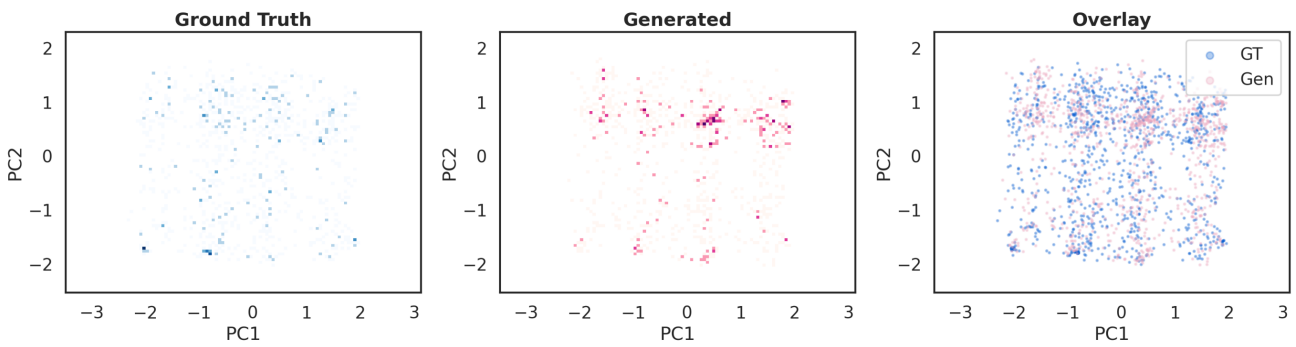

Figure 25. Sequence: dypTwCplhwpa, PCA-CosSim = 0.245, PCA-SW = 0.171

H.3. Conformational Entropy Analysis

We evaluate how well the model captures the conformational diversity of each test sequence by comparing the cluster entropy of generated ensembles against ground truth. Cluster entropy is the Shannon entropy over the full-peptide 12-state conformational code distribution: each conformer is mapped to a string of length N_{res} (one Greek-letter cluster label per residue), and entropy is computed over the frequency distribution of unique codes.

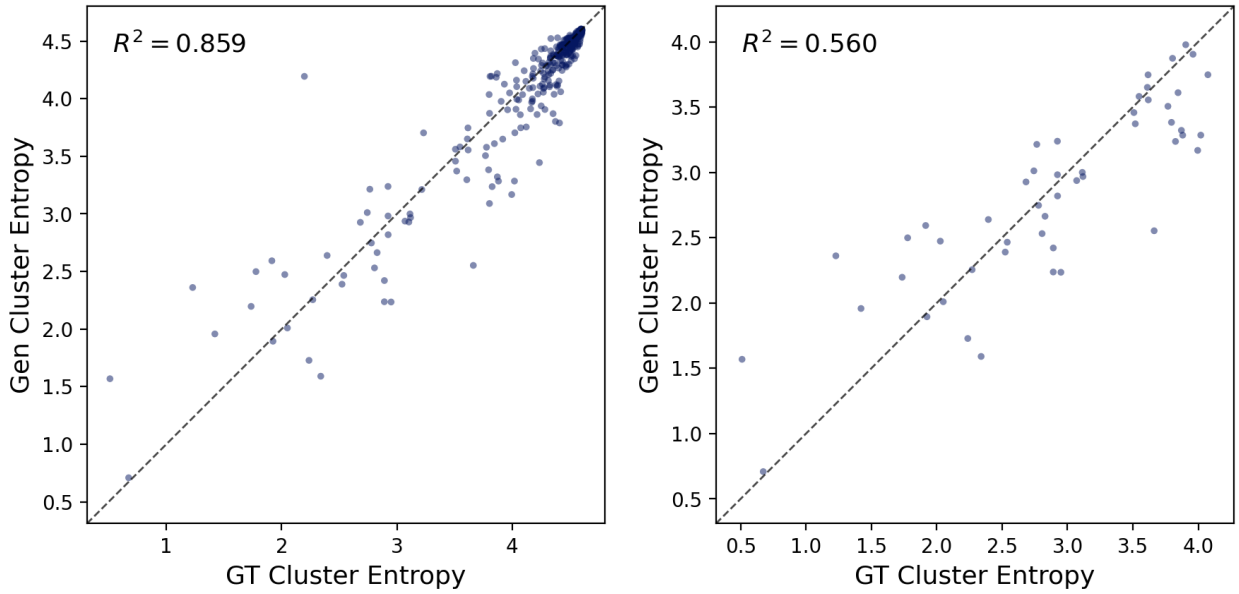

Figure 26. Cluster entropy parity plots for the seq-only test set (Config A). **Left:** All 350 sequences ($R^2 = 0.859$). **Right:** 50 hexapeptides only ($R^2 = 0.560$). Dashed line indicates perfect agreement. Points above the diagonal indicate over-generation of conformational diversity; points below indicate under-generation.

Across all 350 sequences, generated and ground-truth entropies are strongly correlated ($R^2 = 0.859$), demonstrating that the model captures the overall conformational diversity of each peptide. Agreement is tightest for high-entropy sequences (12-mers with many accessible states), where the model consistently generates near-maximum diversity. Hexapeptides show more spread ($R^2 = 0.560$) because their smaller conformational space makes the entropy more sensitive to accurate reproduction of minority cluster populations.

H.3.1. Low-Entropy Cyclic Peptides: Cluster Code Analysis

The most challenging test cases are low-entropy sequences where the ground-truth ensemble is dominated by one or two conformational states. We examine the four lowest-entropy hexapeptides in detail, comparing their full-peptide cluster code

distributions between ground truth and generated ensembles.

Table 17. Cluster code distributions for YImfRP ($H_{\mathrm{GT}} = 0.51$).

Rank	GT Code	GT %	Gen Code	Gen %
1	$\alpha\delta\Gamma\alpha\varepsilon\alpha$	90.5	$\alpha\delta\Gamma\alpha\varepsilon\alpha$	72.1
2	$\varepsilon\gamma\mathrm{B}\Delta\delta\beta$	5.2	$\varepsilon\gamma\mathrm{B}\Delta\delta\beta$	4.3
3	$\alpha\delta\mathrm{AB}\varepsilon\alpha$	0.5	$\alpha\delta\Gamma\alpha\gamma\alpha$	3.1
4	$\alpha\delta\alpha\alpha\gamma\alpha$	0.5	$\varepsilon\gamma\mathrm{B}\Delta\varepsilon\beta$	2.8
5	$\delta\gamma\mathrm{A}\Delta\alpha\beta$	0.4	$\varepsilon\gamma\mathrm{A}\Delta\delta\beta$	1.8

Table 18. Cluster code distributions for lRVPih ($H_{\mathrm{GT}} = 0.67$).

Rank	GT Code	GT %	Gen Code	Gen %
1	$\Gamma\beta\delta\gamma\mathrm{B}\Delta$	85.5	$\Gamma\beta\delta\gamma\mathrm{B}\Delta$	88.7
2	$\Gamma\beta\delta\gamma\mathrm{AE}$	7.7	$\Gamma\beta\delta\gamma\mathrm{AE}$	3.2
3	$\Gamma\beta\delta\zeta\mathrm{AE}$	2.5	$\Gamma\beta\varepsilon\gamma\mathrm{B}\Delta$	0.9
4	$\alpha\zeta\gamma\varepsilon\mathrm{A}\Gamma$	0.8	$\mathrm{B}\delta\gamma\beta\Delta\mathrm{A}$	0.5
5	$\mathrm{AA}\delta\gamma\mathrm{B}\Delta$	0.7	$\mathrm{B}\delta\delta\gamma\Gamma\mathrm{A}$	0.4

Table 19. Cluster code distributions for TanplL ($H_{\mathrm{GT}} = 1.23$).

Rank	GT Code	GT %	Gen Code	Gen %
1	$\gamma\mathrm{BEAB}\varepsilon$	57.5	$\gamma\mathrm{BEAB}\varepsilon$	39.3
2	$\gamma\mathrm{BEAB}\delta$	21.2	$\gamma\mathrm{B}\Delta\mathrm{AB}\varepsilon$	18.2
3	$\gamma\mathrm{BEAZ}\varepsilon$	11.2	$\gamma\mathrm{BEAB}\delta$	8.7
4	$\gamma\mathrm{AEAB}\delta$	2.5	$\gamma\mathrm{BEAZ}\varepsilon$	4.4
5	$\gamma\mathrm{B}\Delta\mathrm{AB}\varepsilon$	1.2	$\gamma\mathrm{B}\Delta\Gamma\alpha\varepsilon$	3.4

Table 20. Cluster code distributions for vlElpW ($H_{\mathrm{GT}} = 1.42$).

Rank	GT Code	GT %	Gen Code	Gen %
1	$\Delta\Gamma\zeta\mathrm{E}\Gamma\beta$	68.0	$\Delta\Gamma\zeta\mathrm{E}\Gamma\beta$	59.0
2	$\Delta\Gamma\beta\Delta\Gamma\beta$	13.4	$\mathrm{E}\Gamma\zeta\mathrm{E}\Gamma\zeta$	7.5
3	$\Delta\mathrm{Z}\beta\Delta\Gamma\beta$	4.1	$\Delta\Gamma\beta\Delta\Gamma\beta$	6.1
4	$\mathrm{AB}\delta\mathrm{AA}\beta$	1.9	$\Delta\Gamma\zeta\Gamma\Gamma\beta$	6.1
5	$\mathrm{E}\Gamma\zeta\mathrm{E}\Gamma\zeta$	1.9	$\Delta\Gamma\zeta\mathrm{E}\Gamma\zeta$	3.3

H.4. Ring Closure Quality

Cyclic peptides require the backbone to form a closed ring via a peptide bond between the C atom of the last residue and the N atom of the first residue. We evaluate ring closure quality by measuring this $\mathrm{C}_{\mathrm{last}}$–$\mathrm{N}_{\mathrm{first}}$ distance for all 350,000 generated conformers (350 sequences $\times$ 1,000 each). An ideal peptide bond has a C–N distance of 1.33 Å; we define conformers with distance > 1.5 Å as "open" (failed ring closure).

Table 21. Ring closure statistics for seq-only generation. % Open = fraction of conformers with $\mathrm{C}_{\mathrm{last}}$–$\mathrm{N}_{\mathrm{first}}$ distance > 1.5 Å.

		ALL	L06	L07	L08	L09	L10	L11	L12
Mean Dist (Å)	Config A	1.340	1.334	1.323	1.326	1.336	1.346	1.361	1.352
	Config B	1.354	1.355	1.340	1.344	1.349	1.357	1.370	1.364
% Open (↓)	Config A	2.72	2.73	1.78	1.87	2.26	2.68	3.92	3.79
	Config B	3.17	3.90	2.27	2.51	2.57	3.13	3.96	3.84

The model achieves proper ring closure (< 1.5 Å) for 97–98% of generated conformers without any explicit geometric constraint during training or inference. Mean closing distances (1.34–1.35 Å) are within 0.01–0.02 Å of the ideal peptide bond length. Config A (optimized temperatures) yields slightly better closure than Config B. Ring closure quality is consistent across peptide lengths, with no degradation for longer sequences (L11–L12 show marginally higher open fractions of $\sim$4%, likely due to increased backbone flexibility rather than systematic failure).

I. Baseline Comparison: Full Results

We evaluate CPGEN against Boltz2 (ensemble generation) and ProteinMPNN (sequence design) on 100 held-out L-amino acid cyclic peptides (50 L06 + 50 L12). Ground truth consists of 1,000 Rosetta-generated conformers per sequence. CPGEN generates 1,000 conformers per sequence using Config A; Boltz2 generates 1,000 conformers under default settings; ProteinMPNN produces 10 sequence designs for each of 100 ground-truth conformers (1,000 total per sequence). All metrics are computed identically to the main evaluation (Section 4.4).

Table 22. Ensemble quality: CPGEN (seq-only) vs. Boltz2 vs. Rosetta GT. Higher is better for RamaCos, ClusterCos, PCA-CosSim; lower is better for PCA-SW. Values are mean ± std over 50 sequences per length.

Comparison	Metric	ALL	L06	L12
CPGEN vs GT	RamaCos ↑	0.509 ± 0.079	0.435 ± 0.038	0.582 ± 0.016
	ClusterCos ↑	0.442 ± 0.442	0.882 ± 0.072	0.003 ± 0.009
	PCA-CosSim ↑	0.379 ± 0.139	0.506 ± 0.072	0.252 ± 0.030
	PCA-SW ↓	0.232 ± 0.126	0.302 ± 0.138	0.163 ± 0.056
Boltz2 vs GT	RamaCos ↑	0.202 ± 0.073	0.159 ± 0.049	0.245 ± 0.068
	ClusterCos ↑	0.207 ± 0.264	0.414 ± 0.232	0.000 ± 0.001
	PCA-CosSim ↑	0.102 ± 0.111	0.170 ± 0.120	0.034 ± 0.036
	PCA-SW ↓	1.406 ± 0.481	1.195 ± 0.431	1.617 ± 0.433
CPGEN vs Boltz2	RamaCos ↑	0.272 ± 0.083	0.280 ± 0.091	0.264 ± 0.074
	ClusterCos ↑	0.223 ± 0.284	0.446 ± 0.249	0.000 ± 0.001
	PCA-CosSim ↑	0.150 ± 0.168	0.267 ± 0.167	0.033 ± 0.032
	PCA-SW ↓	1.397 ± 0.480	1.170 ± 0.396	1.624 ± 0.448

Table 23. Ensemble entropy (Shannon entropy over top-100 cluster codes). Higher indicates greater conformational diversity. Max possible: $\ln(100) \approx 4.61$.

Source	ALL	L06	L12
Rosetta GT	3.774 ± 0.836	2.959 ± 0.259	4.589 ± 0.035
CPGEN	3.542 ± 1.074	2.495 ± 0.339	4.588 ± 0.014
Boltz2	3.445 ± 0.929	2.896 ± 0.982	3.994 ± 0.399

Table 24. Sequence design: CPGEN (rama-only) vs. ProteinMPNN. SeqRec: fraction of predictions exactly matching the ground-truth sequence. PRR: per-residue recovery of the top-1 predicted sequence. Unique: mean number of distinct sequences among 1,000 predictions.

Method	Metric	ALL	L06	L12
CPGEN	SeqRec ↑	0.454 ± 0.459	0.026 ± 0.079	0.882 ± 0.221
	PRR ↑	0.806 ± 0.240	0.617 ± 0.209	0.995 ± 0.020
	Unique	22.6 ± 32.4	40.3 ± 38.2	5.0 ± 3.4
ProteinMPNN	SeqRec ↑	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	PRR ↑	0.121 ± 0.122	0.080 ± 0.112	0.162 ± 0.118
	Unique	597.0 ± 277.5	322.7 ± 48.6	871.3 ± 34.9

CPGEN outperforms both baselines by wide margins across all metrics. For ensemble generation, Boltz2 achieves less than half of CPGEN's RamaCos (0.202 vs. 0.509) and exhibits $6\times$ higher PCA sliced Wasserstein distance (1.406 vs. 0.232), indicating that its predicted conformers occupy a conformational landscape largely disjoint from the Rosetta reference.

Boltz2 also under-generates conformational diversity for 12-mers (entropy 3.99 vs. ground truth4.59), suggesting mode collapse in longer rings. For sequence design, ProteinMPNN achieves zero exact sequence recovery and only 12.1% per-residue recovery—barely above the 5% random baseline for 20 amino acid types—while producing near-maximal sequence diversity (597 unique sequences per 1,000 predictions), indicating that it treats each conformer independently without leveraging ensemble-level information. These results highlight that general-purpose protein structure prediction and design tools, which are trained predominantly on linear proteins from the PDB, do not transfer to the heterochiral cyclic peptide domain without retraining.

J. Parameter Sweep Details

We conduct a systematic hyperparameter sweep on a held-out validation set of 10 sequences of residue-length 6 and residue-length 12, generating 1,000 conformers per sequence. All sweeps use the EMA-averaged checkpoint and report six metrics: Ramachandran cosine similarity (RamaCos), 12-state cluster code cosine similarity (ClusterCos), PCA cosine similarity (PCA-CosSim), PCA sliced Wasserstein distance (PCA-SW), exact sequence recovery (SeqRec), and cluster entropy. Results are reported as the mean over the 10 sequences for each length, as well as the total average. The notation L06 and L12 is used to differentiate the two cyclic peptide lengths (6-mer and 12-mer, respectively)

Guidance weight sweep. We sweep the classifier-free guidance weight $w \in \{0.00, 0.25, 0.50, 0.75, 1.00, 1.50, 2.00, 3.00, 5.00\}$ across all three conditioning modes (seq-only, rama-only, coords-only) with fixed integration steps $= 200$, $T_{\text{CA}} = 0.5$, and $T_{\text{lat}} = 0.1$. Distributional accuracy (RamaCos) peaks at $w = 0.75$ for all three modes: seq-only achieves 0.484, rama-only 0.473, and coords-only 0.525. Beyond $w = 0.75$, quality degrades as over-guidance sharpens the predicted velocity field and collapses ensemble diversity. PCA-SW (measuring global distributional match) is minimized at $w = 0.50$ for seq-only and rama-only, confirming that the optimal guidance weight for distributional fidelity lies below the optimal weight for per-residue accuracy. For seq-only, sequence recovery reaches 99.7% at $w = 0.75$ and 100% at $w \geq 1.0$; rama-only sequence recovery increases monotonically with w (from 0% at $w = 0.0$ to 59.6% at $w = 5.0$), while coords-only never recovers the input sequence regardless of guidance strength.

Fine guidance sweep (rama-only). To resolve the optimal guidance weight for the rama-only mode, we conduct a fine-grained sweep over $w \in \{0.50, 0.55, \ldots, 1.00\}$ (11 values, $\Delta w = 0.05$). RamaCos is largely flat in this range, with a slight peak at $w = 0.75$–0.85 (Avg ≈ 0.477). PCA-SW continues to decrease monotonically as w decreases toward 0.50 (Avg 0.273 at $w = 0.55$ vs. 0.420 at $w = 1.00$), confirming the diversity–accuracy trade-off. Sequence recovery increases steadily from 33.8% ($w = 0.50$) to 57.2% ($w = 1.00$), with L12 peptides reaching 87.5% recovery at $w = 1.00$.

Integration steps sweep. We sweep the number of ODE/SDE integration steps over $\{50, 100, 200, 300, 400\}$ with fixed $w = 1.0$, $T_{\text{CA}} = 0.5$, $T_{\text{lat}} = 0.1$. At 50 steps, all metrics are severely degraded (RamaCos ≈ 0.06 across modes). Quality improves sharply at 100 steps and plateaus at 200 steps, with negligible gains from 300 or 400 steps. For example, seq-only RamaCos is 0.060 (50 steps), 0.330 (100), 0.477 (200), and 0.486 (300). We adopt 200 steps as the default, as it provides near-optimal quality at half the computational cost of 400 steps.

Temperature grid sweep. We sweep a 4×4 grid of SDE noise scaling parameters $T_{\text{CA}} \in \{0.05, 0.1, 0.5, 1.0\}$ and $T_{\text{lat}} \in \{0.05, 0.1, 0.5, 1.0\}$ for all three conditioning modes, with fixed $w = 1.0$ and 200 steps. For RamaCos, the optimal temperature is consistently $T_{\text{CA}} = 0.5$, $T_{\text{lat}} = 0.1$ across modes (seq-only 0.478, rama-only 0.470, coords-only 0.512). Lower T_{CA} values (0.05, 0.1) underperform because insufficient stochastic perturbation limits exploration of the conformational landscape; higher values ($T_{\text{CA}} = 1.0$) add excessive noise that degrades accuracy. The latent channel is less sensitive, but $T_{\text{lat}} = 0.1$ consistently outperforms $T_{\text{lat}} = 0.05$ and larger values. For PCA-SW (diversity), higher temperatures are preferred ($T_{\text{CA}} = 1.0$, $T_{\text{lat}} = 1.0$), achieving the best distributional coverage at the cost of reduced per-residue accuracy. For rama-only sequence recovery, low latent temperatures are critical: $T_{\text{lat}} = 0.05$ yields 58.5% SeqRec (L12: 89.7%), while $T_{\text{lat}} = 1.00$ drops to 48.4%. The coords-only mode is largely insensitive to temperature (PCA-SW varies only 0.182–0.186 across all 16 combinations), consistent with the strong structural constraint imposed by clean C_α coordinates.

J.1. Guidance Weight Sweep

All tables report mean metrics over 10 held-out validation sequences per length, 1,000 generated conformers per sequence, 200 integration steps, T_{CA}=0.5, T_{lat}=0.1 (except where the guidance weight is the swept variable). Bold indicates the best value per mode. ↑ higher is better; ↓ lower is better.

Table 25. RamaCos (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	0.234	0.375	0.465	**0.484**	0.480	0.457	0.415	0.330	0.231
	L06	0.154	0.292	0.383	**0.399**	0.392	0.360	0.305	0.198	0.112
	L12	0.314	0.457	0.547	**0.569**	0.568	0.554	0.524	0.461	0.350
rama_only	Avg	0.235	0.374	0.464	**0.473**	**0.473**	0.462	0.432	0.360	0.264
	L06	0.154	0.287	**0.380**	0.375	0.372	0.362	0.319	0.223	0.132
	L12	0.315	0.461	0.549	0.570	**0.573**	0.562	0.545	0.498	0.397
coords_only	Avg	0.224	0.445	0.513	**0.525**	0.513	0.459	0.339	0.178	0.092
	L06	0.143	0.386	0.472	**0.492**	0.481	0.415	0.263	0.136	0.070
	L12	0.304	0.504	0.554	**0.557**	0.544	0.502	0.415	0.220	0.115

Table 26. ClusterCos (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	0.035	0.210	0.347	**0.353**	0.338	0.317	0.276	0.178	0.078
	L06	0.071	0.419	0.694	**0.704**	0.673	0.629	0.549	0.353	0.156
	L12	0.000	0.000	0.001	0.002	0.002	**0.004**	0.003	0.002	0.000
rama_only	Avg	0.036	0.220	0.334	**0.338**	0.333	0.326	0.303	0.224	0.112
	L06	0.072	0.441	0.666	**0.675**	0.664	0.649	0.603	0.444	0.222
	L12	0.000	0.000	0.001	0.002	**0.003**	**0.003**	**0.003**	**0.003**	0.001
coords_only	Avg	0.029	0.335	0.445	**0.458**	**0.458**	0.428	0.325	0.146	0.033
	L06	0.058	0.653	0.818	0.829	**0.833**	0.803	0.635	0.291	0.066
	L12	0.000	0.018	0.073	**0.087**	0.082	0.054	0.014	0.000	0.000

Table 27. PCA-CosSim (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	0.044	0.225	**0.327**	0.315	0.297	0.279	0.230	0.159	0.124
	L06	0.074	0.261	**0.393**	0.377	0.352	0.328	0.224	0.105	0.051
	L12	0.014	0.189	**0.260**	0.252	0.242	0.230	0.236	0.213	0.196
rama_only	Avg	0.047	0.231	**0.314**	0.301	0.291	0.266	0.246	0.198	0.134
	L06	0.081	0.265	**0.360**	0.343	0.343	0.303	0.263	0.188	0.078
	L12	0.013	0.196	**0.267**	0.259	0.239	0.228	0.230	0.209	0.191
coords_only	Avg	0.030	0.333	0.375	0.376	**0.379**	0.325	0.259	0.115	0.019
	L06	0.053	0.405	0.510	0.512	**0.519**	0.400	0.251	0.103	0.023
	L12	0.006	0.261	0.241	0.239	0.239	0.250	**0.266**	0.128	0.014

Table 28. PCA-SW (↓) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	1.559	0.665	**0.247**	0.315	0.383	0.472	0.548	0.702	0.917
	L06	1.175	0.533	**0.367**	0.473	0.573	0.679	0.782	0.958	1.147
	L12	1.943	0.798	**0.127**	0.157	0.193	0.264	0.314	0.447	0.688
rama_only	Avg	1.594	0.686	**0.272**	0.368	0.419	0.500	0.555	0.663	0.842
	L06	1.239	0.541	**0.411**	0.578	0.633	0.727	0.764	0.864	1.096
	L12	1.950	0.832	**0.132**	0.158	0.206	0.273	0.346	0.462	0.589
coords_only	Avg	1.698	0.236	0.191	0.188	**0.185**	0.190	0.223	0.769	1.409
	L06	1.266	0.278	0.212	0.208	**0.207**	0.222	0.286	0.488	1.011
	L12	2.129	0.193	0.171	0.168	0.163	**0.158**	0.160	1.050	1.808

Table 29. SeqRec (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	0.000	0.000	0.677	0.997	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
	L06	0.000	0.001	0.769	0.996	0.999	**1.000**	**1.000**	**1.000**	**1.000**
	L12	0.000	0.000	0.585	0.998	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
rama_only	Avg	0.000	0.000	0.344	0.561	0.568	0.579	0.586	0.588	**0.596**
	L06	0.000	0.000	0.205	0.260	0.261	0.267	**0.274**	0.273	0.273
	L12	0.000	0.000	0.482	0.863	0.874	0.891	0.899	0.903	**0.920**
coords_only	Avg	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	L06	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	L12	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Table 30. Entropy (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	**4.485**	4.349	3.975	3.801	3.765	3.836	3.996	4.155	4.249
	L06	**4.364**	4.093	3.352	3.013	2.964	3.127	3.439	3.763	3.934
	L12	**4.605**	**4.605**	4.598	4.589	4.566	4.545	4.553	4.547	4.564
rama_only	Avg	**4.495**	4.325	3.935	3.741	3.717	3.754	3.871	4.071	4.163
	L06	**4.384**	4.044	3.269	2.899	2.865	2.964	3.214	3.609	3.772
	L12	**4.605**	**4.605**	4.600	4.583	4.568	4.544	4.528	4.534	4.554
coords_only	Avg	4.503	4.160	3.973	3.954	4.016	4.161	4.315	4.494	**4.553**
	L06	4.401	3.822	3.423	3.373	3.500	3.796	4.119	4.384	**4.500**
	L12	**4.605**	4.498	4.524	4.535	4.531	4.526	4.511	4.604	**4.605**

Table 31. SeqUnique (↑) — Guidance weight sweep.

Mode	Length	0.00	0.25	0.50	0.75	1.00	1.50	2.00	3.00	5.00
seq_only	Avg	**999.2**	978.9	112.6	2.4	1.1	1.1	1.0	1.1	1.2
	L06	**998.3**	957.8	54.3	2.1	1.1	1.2	1.0	1.0	1.0
	L12	**1000.0**	**1000.0**	170.9	2.6	1.0	1.0	1.0	1.1	1.3
rama_only	Avg	**999.5**	987.9	182.7	31.5	25.7	23.4	21.9	21.7	20.1
	L06	**998.9**	975.8	169.0	57.1	47.7	43.4	40.3	40.2	36.6
	L12	**1000.0**	**1000.0**	196.3	5.8	3.6	3.4	3.5	3.2	3.6
coords_only	Avg	**999.3**	998.5	997.4	996.2	995.0	991.6	988.0	993.3	999.0
	L06	**998.5**	996.9	994.8	992.4	990.0	983.2	976.0	986.7	997.9
	L12	**1000.0**	**1000.0**	**1000.0**	**1000.0**	**1000.0**	**1000.0**	999.9	999.9	**1000.0**

Table 32. RamaCos (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	0.465	0.471	0.472	0.476	0.471	**0.477**	0.473	**0.477**	0.474	0.467	0.474
L06	0.375	**0.383**	0.379	**0.383**	0.376	0.379	0.374	0.378	0.375	0.366	0.371
L12	0.554	0.559	0.564	0.569	0.566	0.575	0.572	0.575	0.572	0.568	**0.576**

Table 33. ClusterCos (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	0.334	0.336	0.333	**0.344**	0.343	**0.344**	0.336	0.342	0.335	0.335	0.331
L06	0.667	0.671	0.666	**0.687**	0.685	0.686	0.670	0.681	0.668	0.667	0.659
L12	0.001	0.001	0.001	0.002	0.002	0.002	**0.003**	0.002	0.002	**0.003**	**0.003**

Table 34. PCA-CosSim (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	0.314	**0.323**	0.310	0.315	0.309	0.303	0.300	0.289	0.283	0.284	0.290
L06	0.369	0.382	0.362	**0.391**	0.378	0.369	0.351	0.343	0.319	0.329	0.332
L12	0.260	**0.265**	0.259	0.239	0.241	0.237	0.248	0.235	0.247	0.238	0.247

Table 35. PCA-SW (↓) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	0.283	**0.273**	0.303	0.316	0.332	0.350	0.380	0.388	0.386	0.404	0.420
L06	**0.424**	0.429	0.472	0.489	0.526	0.550	0.587	0.599	0.601	0.635	0.658
L12	0.141	**0.117**	0.133	0.142	0.138	0.150	0.172	0.178	0.171	0.174	0.182

Table 36. SeqRec (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	0.338	0.474	0.529	0.545	0.554	0.559	0.561	0.563	0.565	0.567	**0.572**
L06	0.205	0.242	0.249	0.254	0.258	0.261	0.258	0.262	0.264	0.265	**0.269**
L12	0.470	0.706	0.809	0.837	0.850	0.857	0.865	0.863	0.865	0.869	**0.875**

Table 37. Entropy (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	**3.940**	3.874	3.830	3.800	3.764	3.761	3.748	3.739	3.732	3.717	3.716
L06	**3.279**	3.147	3.068	3.009	2.939	2.931	2.909	2.902	2.893	2.861	2.852
L12	**4.601**	**4.601**	4.593	4.592	4.588	4.591	4.587	4.575	4.571	4.573	4.581

Table 38. SeqUnique (↑) — Fine guidance sweep (rama_only, $w = 0.50$–1.00).

Length	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
Avg	**182.7**	86.2	50.6	38.0	33.6	30.1	29.3	29.2	27.3	26.1	25.7
L06	**173.9**	104.0	74.1	63.3	58.9	54.5	53.7	53.8	50.7	48.2	47.7
L12	**191.4**	68.3	27.0	12.6	8.3	5.7	4.8	4.6	3.8	4.0	3.6

J.2. Integration Steps Sweep

All tables: w=1.0, T_{CA}=0.5, T_{lat}=0.1, 10 val seqs/length, 1,000 samples. Quality plateaus at 200 steps.

Table 39. RamaCos (↑) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	0.060	0.330	0.477	**0.486**	0.485
	L06	0.043	0.271	0.389	**0.398**	0.392
	L12	0.078	0.390	0.564	0.575	**0.579**
rama_only	Avg	0.060	0.326	0.474	0.476	**0.478**
	L06	0.041	0.260	**0.376**	0.374	0.373
	L12	0.079	0.392	0.573	0.577	**0.583**
coords_only	Avg	0.065	0.253	0.512	**0.514**	0.513
	L06	0.041	0.221	0.480	**0.483**	0.480
	L12	0.090	0.285	0.544	**0.545**	**0.545**

Table 40. ClusterCos (↑) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	0.001	0.275	0.340	**0.341**	0.340
	L06	0.002	0.549	0.678	**0.680**	0.678
	L12	0.000	0.000	0.002	**0.003**	**0.003**
rama_only	Avg	0.002	0.278	0.333	**0.334**	0.333
	L06	0.004	0.555	0.663	**0.665**	0.662
	L12	0.000	0.000	0.002	**0.004**	**0.004**
coords_only	Avg	0.002	0.293	0.457	**0.460**	**0.460**
	L06	0.004	0.586	0.830	**0.836**	0.835
	L12	0.000	0.001	**0.085**	0.084	**0.085**

Table 41. PCA-CosSim (↑) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	0.006	0.234	0.296	0.304	**0.305**
	L06	0.009	0.220	0.360	**0.373**	0.370
	L12	0.004	**0.247**	0.232	0.235	0.239
rama_only	Avg	0.007	0.233	0.291	0.287	**0.293**
	L06	0.008	0.216	0.335	0.333	**0.347**
	L12	0.005	**0.250**	0.246	0.240	0.240
coords_only	Avg	0.011	0.242	**0.381**	0.371	0.372
	L06	0.013	0.212	**0.516**	0.507	0.499
	L12	0.009	**0.271**	0.247	0.235	0.244

Table 42. PCA-SW (↓) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	1.978	**0.319**	0.388	0.380	0.386
	L06	1.802	**0.477**	0.579	0.567	0.580
	L12	2.153	**0.160**	0.197	0.193	0.192
rama_only	Avg	1.996	**0.344**	0.410	0.432	0.414
	L06	1.833	**0.514**	0.628	0.662	0.636
	L12	2.159	**0.175**	0.193	0.201	0.191
coords_only	Avg	1.719	0.251	0.185	0.186	**0.184**
	L06	1.507	0.294	**0.207**	0.208	**0.207**
	L12	1.930	0.208	0.162	0.163	**0.161**

Table 43. SeqRec (↑) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	0.968	**0.999**	**0.999**	**0.999**	**0.999**
	L06	0.972	**0.999**	**0.999**	**0.999**	**0.999**
	L12	0.965	0.999	**1.000**	**1.000**	**1.000**
rama_only	Avg	0.548	**0.572**	0.570	0.569	0.565
	L06	0.259	**0.270**	0.265	0.264	0.260
	L12	0.836	**0.875**	**0.875**	0.873	0.871
coords_only	Avg	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	L06	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
	L12	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**

Table 44. Entropy (↑) — Integration steps sweep.

Mode	Length	50	100	200	300	400
seq_only	Avg	**4.586**	4.338	3.762	3.733	3.706
	L06	**4.567**	4.071	2.953	2.905	2.853
	L12	**4.605**	**4.605**	4.571	4.562	4.559
rama_only	Avg	**4.584**	4.359	3.704	3.685	3.672
	L06	**4.564**	4.113	2.837	2.807	2.776
	L12	**4.605**	**4.605**	4.570	4.564	4.568
coords_only	Avg	**4.591**	4.421	4.016	4.020	4.022
	L06	**4.577**	4.240	3.494	3.503	3.503
	L12	**4.605**	4.603	4.538	4.536	4.541

J.3. Temperature Grid Sweep

4×4 grid over C_α SDE noise scale ($T_{\text{CA}} \in \{0.05, 0.1, 0.5, 1.0\}$) and latent SDE noise scale ($T_{\text{lat}} \in \{0.05, 0.1, 0.5, 1.0\}$). All runs: w=1.0, 200 steps, 10 val seqs/length, 1,000 samples. Columns are labeled by (T_{CA}, T_{lat}) pairs.

Table 45. RamaCos (↑) — Temperature grid, seq_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.411	0.422	0.456	0.457	0.438	0.444	0.471	0.462	0.472	0.478	**0.482**	0.451	0.460	0.464	0.447	0.413
L06	0.301	0.313	0.369	0.378	0.330	0.341	0.385	0.383	0.379	0.392	**0.411**	0.384	0.385	0.393	0.385	0.356
L12	0.521	0.530	0.543	0.535	0.545	0.548	0.556	0.541	**0.565**	**0.565**	0.553	0.517	0.535	0.535	0.508	0.469

Table 46. ClusterCos (↑) — Temperature grid, seq_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.302	0.305	0.332	0.345	0.311	0.313	0.338	0.353	0.333	0.339	0.357	0.357	0.343	0.346	**0.359**	0.351
L06	0.598	0.604	0.658	0.688	0.614	0.619	0.672	0.703	0.663	0.673	0.713	0.714	0.685	0.692	**0.718**	0.702
L12	0.007	0.006	0.005	0.002	**0.008**	0.006	0.004	0.003	0.003	0.004	0.001	0.001	0.001	0.001	0.000	0.000

Table 47. PCA-SW (↓) — Temperature grid, seq_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.812	0.771	0.624	0.507	0.702	0.670	0.547	0.451	0.387	0.379	0.320	0.264	0.261	0.243	0.207	**0.184**
L06	0.988	0.948	0.794	0.649	0.902	0.876	0.722	0.597	0.572	0.570	0.470	0.393	0.395	0.360	0.305	**0.284**
L12	0.635	0.594	0.454	0.365	0.501	0.463	0.372	0.305	0.202	0.187	0.171	0.135	0.126	0.127	0.109	**0.085**

Table 48. SeqRec (↑) — Temperature grid, seq_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.999	0.999	0.995	0.973	0.999	0.999	0.996	0.975	**1.000**	0.999	0.995	0.974	**1.000**	**1.000**	0.995	0.969
L06	0.998	0.998	0.993	0.980	0.998	0.998	0.995	0.981	0.999	0.998	0.995	0.983	0.999	**1.000**	0.995	0.980
L12	**1.000**	**1.000**	0.997	0.967	**1.000**	**1.000**	0.997	0.968	**1.000**	**1.000**	0.996	0.966	**1.000**	**1.000**	0.996	0.958

Table 49. RamaCos (↑) — Temperature grid, rama_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.395	0.407	0.449	0.452	0.428	0.432	0.459	0.461	0.469	**0.470**	**0.470**	0.445	0.462	0.461	0.449	0.411
L06	0.260	0.278	0.344	0.366	0.302	0.308	0.358	0.380	0.364	0.367	**0.387**	0.373	0.381	0.383	0.384	0.356
L12	0.530	0.536	0.553	0.539	0.554	0.556	0.560	0.542	**0.574**	**0.574**	0.552	0.517	0.543	0.538	0.513	0.466

Table 50. ClusterCos (↑) — Temperature grid, rama_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.279	0.286	0.326	0.346	0.294	0.302	0.331	0.352	0.327	0.335	0.350	**0.361**	0.344	0.345	0.358	0.346
L06	0.550	0.565	0.647	0.688	0.581	0.597	0.657	0.701	0.652	0.666	0.698	**0.721**	0.686	0.689	0.715	0.692
L12	**0.008**	**0.008**	0.006	0.004	0.007	0.007	0.005	0.003	0.002	0.004	0.002	0.001	0.001	0.001	0.000	0.000

Table 51. PCA-SW (↓) — Temperature grid, rama_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.890	0.867	0.694	0.565	0.766	0.740	0.621	0.504	0.431	0.417	0.351	0.276	0.269	0.258	0.216	**0.194**
L06	1.137	1.119	0.912	0.769	1.027	1.009	0.847	0.710	0.655	0.646	0.531	0.432	0.422	0.409	0.341	**0.300**
L12	0.642	0.616	0.475	0.361	0.505	0.471	0.395	0.298	0.208	0.188	0.171	0.120	0.116	0.107	0.091	**0.088**

Table 52. SeqRec (↑) — Temperature grid, rama_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	**0.585**	0.579	0.557	0.502	0.579	0.573	0.551	0.492	0.573	0.567	0.545	0.491	0.569	0.566	0.547	0.484
L06	**0.273**	0.265	0.255	0.225	0.269	0.267	0.251	0.227	0.264	0.264	0.249	0.219	0.265	0.264	0.249	0.220
L12	**0.897**	0.892	0.860	0.779	0.888	0.879	0.852	0.757	0.883	0.869	0.841	0.762	0.873	0.869	0.844	0.747

Table 53. RamaCos (↑) — Temperature grid, coords_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.497	0.510	0.501	0.464	0.500	0.510	0.498	0.466	0.501	**0.512**	0.498	0.458	0.503	0.509	0.485	0.445
L06	0.470	0.475	0.462	0.419	0.474	**0.480**	0.457	0.421	0.471	0.477	0.459	0.417	0.472	0.474	0.445	0.400
L12	0.525	0.545	0.541	0.509	0.526	0.541	0.540	0.511	0.532	**0.547**	0.536	0.500	0.535	0.543	0.526	0.489

Table 54. ClusterCos (↑) — Temperature grid, coords_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.454	0.457	0.450	0.427	0.458	0.456	0.448	0.425	0.455	**0.460**	0.446	0.421	0.455	0.456	0.441	0.415
L06	0.826	0.831	0.824	0.799	0.832	0.830	0.819	0.795	0.829	**0.838**	0.821	0.791	0.833	0.836	0.820	0.787
L12	0.081	**0.083**	0.075	0.054	**0.083**	**0.083**	0.077	0.055	0.081	0.081	0.072	0.052	0.077	0.075	0.061	0.043

Table 55. PCA-SW (↓) — Temperature grid, coords_only.

	T_{CA}=0.05				T_{CA}=0.10				T_{CA}=0.50				T_{CA}=1.00			
T_{lat}	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00	0.05	0.10	0.50	1.00
Avg	0.186	0.186	0.184	0.183	0.186	0.185	0.183	0.184	0.185	0.184	**0.182**	0.186	0.185	0.184	0.183	**0.182**
L06	0.208	0.209	0.212	0.217	**0.207**	0.208	0.210	0.218	0.209	**0.207**	0.212	0.222	0.209	0.209	0.214	0.222
L12	0.164	0.164	0.156	0.149	0.164	0.162	0.156	0.150	0.161	0.161	0.151	0.149	0.161	0.159	0.152	**0.142**