CYCLOFLOW: INTERNAL COORDINATE FLOW MATCHING AND GEOMETRIC RING CLOSURE FOR CYCLIC PEPTIDE CONFORMER GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

A cyclic peptide is a short peptide whose backbone is closed into a ring. Cyclic peptides can engage broad protein interfaces while remaining cell membrane permeable. Both properties depend on a conformational ensemble rather than on a single structure, so the conformational ensemble is what has to be predicted. A generative model of that ensemble has to choose the coordinates it works in. Internal coordinates suit a cyclic peptide. Ring closure ties the torsion angles to one another, and the representation encodes that dependency for free where a Cartesian one has to learn it. We present CYCLOFLOW, which transports a bond length, a bond angle and a torsion angle at every heavy atom by flow matching, each from a prior chosen for that internal coordinate, and closes the ring by solving six of these internal coordinates. On CREMP tetra-, penta- and hexapeptides, CYCLOFLOW leads the strongest baseline by 4.8 and 10.4 points of ring and all-atom coverage, reduces the Wasserstein distance to the reference ensembles, and clashes in only 1.4% of its conformers. It follows the reference distribution of each internal coordinate closely, and places pairs of ring torsion angles, holding more of their coupling. The code is available at `https://anonymous.4open.science/r/cycloflow`.

1 INTRODUCTION

A cyclic peptide is a peptide whose residues are joined into a ring. Cyclic peptides are attractive therapeutics. They present enough surface to engage the broad, flat interfaces that small molecules bind poorly, and the ring protects them from proteolytic degradation (Driggers et al., 2008; Vinogradov et al., 2019). Many are orally bioavailable despite being larger and more polar than oral small molecules (Lipinski et al., 2001; Nielsen et al., 2017). Some act as molecular chameleons, polar in water and folded in a membrane with backbone polarity shielded by intramolecular hydrogen bonds (Nielsen et al., 2017; Wang et al., 2018). Both behaviors depend on which conformers a peptide populates. The chameleon shifts that balance with its surroundings, and binding needs the conformer the target takes to be one the peptide already occupies (Research et al., 2026). A generator therefore needs to reproduce both the conformers in the ensemble and their relative populations (Grambow et al., 2024).

Reference ensembles come from high-quality but computationally expensive physics-based conformational search (Grambow et al., 2024). Deep-learning generators produce those ensembles in its place, and each has to choose the coordinates it works in. Internal coordinates describe a structure by the bond lengths, bond angles and torsion angles between bonded atoms rather than by the positions of the atoms themselves, so they hold the torsion angles that carry a cyclic peptide's flexibility directly, where a Cartesian model has to recover them from atom positions. Vilya-1 works in Cartesian coordinates, diffusing over atom positions, so one representation serves canonical and noncanonical residues and no chemistry has to be enumerated in advance (Research et al., 2026). RINGER works in internal coordinates, diffusing over the backbone bond angles and torsion angles and taking bond lengths from training set means (Grambow et al., 2023). It gives its bond angles and torsion angles the same diffusion noise schedule, although a bond angle moves far less than a torsion angle, and it closes the ring afterwards by optimizing all ring atom positions in Cartesian space.

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

Here we present CYCLOFLOW, a model that generates cyclic peptide ensembles in full internal coordinates by flow matching. Every bond length, bond angle and torsion angle at every heavy atom comes from the model. Each type of internal coordinate travels its own path from its own prior, rather than one transport serving all three. This is achieved by Diffusion Transformer (DiT) blocks (Peebles & Xie, 2023a) over atom-level internal coordinates, which carries the velocity of every internal coordinate at once, so the ring and its side chains are generated jointly. The ring is then closed by solving six internal coordinates, while leaving every other coordinate as flow matching proposed it. The all-atom structure is rebuilt in Cartesian coordinates from these internal coordinates, and a steric refinement relieves the close contacts that remain. CYCLOFLOW outperforms all generators we compare against on coverage, on matching and on the Wasserstein distance to the reference ensembles.

Our contributions are summarized as follows:

- **Every internal coordinate is generated.** One conformer differs from another mostly in its torsion angles, so those are what a generator has to vary. In a ring they are not free to vary one at a time, since turning one torsion swings every atom after it and the ring has to close. Internal coordinates encode that dependence. Bond lengths and bond angles move far less, but they do move, and restoring them from training set means gives every conformer the same value, so we transport all three at every heavy atom.
- **Each internal coordinate is transported on its own terms.** The three move on scales orders of magnitude apart, a median 0.0018 Å for a bond length against $0.63°$ for a bond angle and $8.61°$ for a free torsion angle. Rescaling absorbs that gap, but not the periodicity of an angle. Bond lengths and bond angles start from Gaussians at their own scales, and torsion angles from the uniform distribution on the circle, traveling the wrapped geodesic.
- **Closure is solved on six coordinates.** Closing the ring is a condition on a rigid frame in three dimensions, which has six degrees of freedom. Solving exactly those six leaves the rest of the molecule as flow matching proposed it. Removing the solve breaks the ring while the rest changes little, and replacing it with optimization over all ring atom positions moves the rest of the molecule with the ring.

2 RELATED WORK

Physics-based conformer sampling. Cyclic peptide ensembles are usually built by search under a physical energy model, which becomes expensive as the conformational space grows. CREST (Pracht et al., 2020) pairs metadynamics with a semi-empirical energy, and the CREMP ensembles we use as references come from CREST with GFN2-xTB (Bannwarth et al., 2019; Grambow et al., 2024). RDKit ETKDG is the cheap alternative, combining distance geometry with torsional preferences and ring-specific heuristics (Wang et al., 2020), usually minimized under MMFF94 (Halgren, 1996).

Learning-based conformer generation. Most learned generators were developed on small molecules. Four of the generators we compare against work in Cartesian coordinates, MCF-S (Wang et al., 2024), ET-Flow-*SO(3)* (Hassan et al., 2024), AGDiff (Vieira Wyzykowski et al., 2025) and $\text{AvgFlow}_{\text{DiT}}$ (Cao et al., 2025). Torsional Diffusion (Jing et al., 2022) works in internal coordinates and diffuses only over freely rotatable torsion angles, taking the ring conformation from RDKit. Using RDKit for ring conformations may suffice for rigid drug-like rings, but leaves the central task of cyclic peptide backbone sampling to RDKit.

Cyclic peptides. RINGER is the closest prior work and the one we compare against throughout. It diffuses over the bond angles and torsion angles of the ring, giving both the same diffusion noise schedule although they move orders of magnitude apart, and restores bond lengths from training set means (Grambow et al., 2023). Each ring atom is conditioned on a Morgan fingerprint of its side chain, and the ring is closed afterwards by an SLSQP optimization over all ring atom positions. Vilya-1 covers canonical and noncanonical residues in one all-atom Cartesian representation (Research et al., 2026), and pays for that generality by predicting the local geometry internal coordinates hold nearly fixed. We do not report it as a baseline, since no implementation or weights are public.

Appendix A places every method named here by what it produces and what transports it.

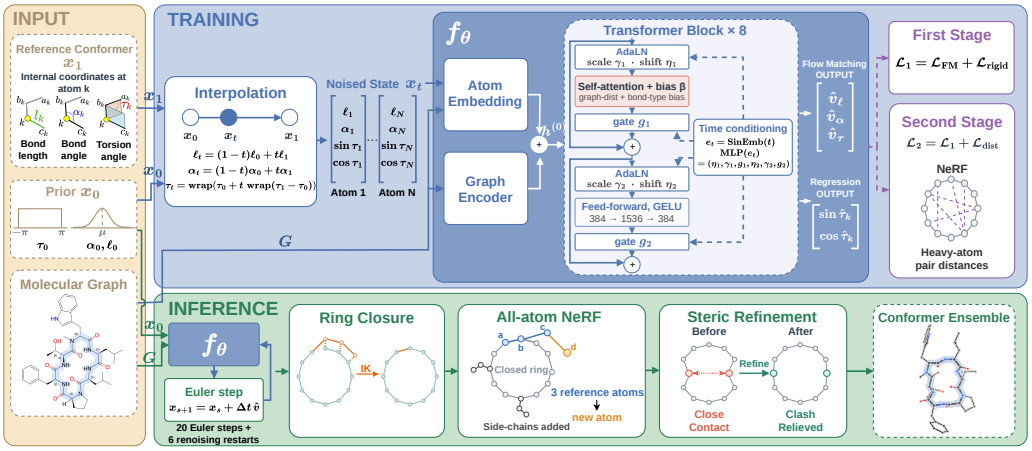

Figure 1: The CYCLOFLOW architecture, in three parts, the inputs, training and inference. CYCLOFLOW generates the internal coordinates of every heavy atom by flow matching, with a Transformer of DiT blocks carrying the velocity. Training runs in two stages. At inference the ring is closed by solving six internal coordinate degrees of freedom, a NeRF pass then rebuilds the all-atom structure, and a short steric refinement relieves the contacts that remain.

3 METHOD

CYCLOFLOW generates the internal coordinates of a conformer and then closes the ring on them (Figure 1). Two choices decide the design, what to generate and how to transport it. We generate a bond length, a bond angle and a torsion angle at every heavy atom. Flow matching prescribes the path rather than deriving it from a corruption process, so each of the three internal coordinates has a prior and a path of its own. Appendix B measures the claims behind both choices. Appendix C gives every setting this section leaves out and Appendix D the train and inference cost.

3.1 INPUT REPRESENTATION

Molecular features. A residue sequence is expanded into an RDKit molecular graph. Every atom k carries an atomic number Z_k and a feature vector f_k holding its degree, aromaticity, ring membership, hydrogen count and a chirality code assigned by the Cahn–Ingold–Prelog rules (Cahn et al., 1966). A flag r_k marks whether its torsion is rotatable, turning freely about the bond it is measured on. The torsion angles with $r_k = 0$ are the *rigid* ones, held near a fixed value by a planar amide bond, by unsaturation, or by a small ring fused onto the backbone as in proline. A flag s_k marks whether the atom holds one of the ring closure coordinates, and we keep its input slot but hold it at zero. Appendix C.2 gives the rule that sets r_k and the reason s_k is hidden. Each bond has a four-dimensional feature indicating single, multiple, aromatic and in-ring bonds.

Internal coordinates. Atoms are ordered ring first, side chains outward. The order fixes a chain a_k, b_k, k, c_k through each atom k, which has three internal coordinates measured on that chain,

$$\ell_k = \|p_k - p_{c_k}\|, \qquad \alpha_k = \angle(p_{b_k}, p_k, p_{c_k}), \qquad \tau_k = \mathrm{dih}(p_{a_k}, p_{b_k}, p_k, p_{c_k}), \tag{1}$$

with $p_k \in \mathbb{R}^3$ the Cartesian position of atom k. Here ℓ_k is the k–c_k bond length, α_k the bond angle with its vertex at k, and τ_k the torsion angle about b_k–k, the middle bond of the four-atom chain, drawn on one atom in the input panel of Figure 1. These three are what flow matching transports. A training example gives x_1, the internal coordinates of a reference conformer, and the prior gives x_0. During training the network sees the interpolant between them at flow time t (Section 3.3). At inference x_1 is not available and the trajectory is integrated from x_0 alone. Either way the three coordinates are measured afresh at every step.

Graph encoder. A graph encoder $\mathrm{GE}(\cdot)$ learns a representation of the molecular graph. It sees G with the four-dimensional bond feature on each edge, and on each atom the feature vector f_k,

the flags r_k and s_k, and a one-hot composite type crossing Z_k with the aromatic and in-ring flags and with whether the atom lies on the ring or a side chain. The internal coordinates are left out because they change with t, so $\mathrm{GE}(G)$ is computed once per molecule and stays constant along the trajectory. Atoms that constrain one another sit far apart in the graph of a cyclic peptide, so we use GraphCliff (Kim et al., 2026), which widens the receptive field within a layer rather than by depth.

Atom embedding. The molecular features and the internal coordinates meet here,

$$h_k^{(0)} = \mathrm{Emb}(Z_k) + W_{\mathrm{node}} \left[f_k,\ r_k,\ s_k,\ \ell_k^{(t)},\ \alpha_k^{(t)},\ \sin\tau_k^{(t)},\ \cos\tau_k^{(t)} \right] + W_{\mathrm{ge}}\,\mathrm{GE}(G)_k, \quad (2)$$

where the superscript on h_k indexes the Transformer layer, $\mathrm{Emb}(\cdot) \in \mathbb{R}^{d_{\mathrm{model}}}$ is a learned embedding table indexed by atomic number, and W_{node} projects the concatenated features to the same hidden width d_{model}. Torsion angles enter as $(\sin\tau, \cos\tau)$ to preserve periodicity.

3.2 Velocity field

Internal coordinates are already invariant to rigid motion, so the velocity field needs no equivariance. We parameterize it with an atom-level Transformer of eight DiT blocks (Peebles & Xie, 2023b).

Time conditioning. The flow time is one scalar per molecule. A sinusoidal embedding turns it into a vector e_t, and each block ℓ has its own MLP that reads that vector and emits the six modulation parameters of its two sublayers,

$$\left(\eta_1^\ell, \gamma_1^\ell, g_1^\ell,\ \eta_2^\ell, \gamma_2^\ell, g_2^\ell\right)(t) = \mathrm{MLP}_\ell(e_t)\,, \quad (3)$$

a shift η^ℓ, a scale γ^ℓ and a gate g^ℓ for the attention sublayer and the same three for the feed-forward one. Every block sees the same e_t, so what differs between them is MLP_ℓ.

Transformer block. Write h for the stack of atom embeddings entering block ℓ, with $h^{(0)}$ given by equation (2). The attention sublayer runs first and the feed-forward sublayer takes its output. Each normalizes its input, applies the shift and scale, runs its own operation, and passes the result through its gate before the residual add,

$$h' = h + g_1^\ell\,\mathrm{Attn}\big(\mathrm{LN}_1(h)\odot(1+\gamma_1^\ell)+\eta_1^\ell\big)\,, \quad h'' = h' + g_2^\ell\,\mathrm{FF}\big(\mathrm{LN}_2(h')\odot(1+\gamma_2^\ell)+\eta_2^\ell\big)\,. \quad (4)$$

Both normalizations are parameter-free, the affine part coming entirely from (η^ℓ, γ^ℓ), so the flow time reaches a block only through the six modulation parameters of equation (3). Attention runs over every atom and on its own knows nothing about connectivity. We add to each query–key logit a learned per-head bias derived from the pair's graph shortest-path distance and its bond type. On a cyclic peptide that distance follows the ring rather than a linear sequence, so the two atoms on either side of the closing bond are recognized as neighbors and the ring has no break point.

Output heads. Three MLP heads take each atom's final hidden state and emit one velocity each, for the bond length, the bond angle and the torsion angle. A separate head predicts $(\sin\tau, \cos\tau)$ for the rigid torsion angles.

3.3 Training

We train by flow matching over internal coordinates. A training example is a single reference conformer rather than a molecule, and x_1 holds its internal coordinates. At time $t \in [0,1]$ the model sees the interpolant between x_1 and a prior sample x_0 and predicts the velocity toward x_1. Each coordinate is drawn from its own prior, and its target velocity is the difference between the reference value and that draw,

$$\begin{aligned} \ell^{(0)} &\sim \mathcal{N}(\mu_\ell, \sigma_\ell^2), & \alpha^{(0)} &\sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), & \tau^{(0)} &\sim \mathcal{U}(-\pi, \pi), \\ v_\ell &= \ell^{(1)} - \ell^{(0)}, & v_\alpha &= \alpha^{(1)} - \alpha^{(0)}, & \Delta_\tau &= \mathrm{wrap}\Big(\tau^{(1)} - \tau^{(0)}\Big)\,. \end{aligned} \quad (5)$$

Bond lengths and bond angles interpolate linearly, so at time t the model sees $\ell_t = (1-t)\ell^{(0)} + t\,\ell^{(1)}$ and α_t likewise. A torsion angle lives on the circle and follows the wrapped geodesic instead,

$\tau_t = \mathrm{wrap}(\tau^{(0)} + t\,\Delta_\tau)$, with $\mathrm{wrap}(\cdot)$ mapping angles back to $[-\pi, \pi)$ so it takes the shorter of the two arcs between its endpoints. The prior means and widths are generic chemistry values rather than dataset statistics, so the sampler needs nothing about a molecule beyond G.

Training runs in two stages. The output heads emit velocities $\hat{v}_\ell, \hat{v}_\alpha, \hat{v}_\tau$ at the interpolated state x_t, and a separate head emits $\hat{q}_k \in \mathbb{R}^2$ at each rigid torsion angle. The first stage combines the flow-matching loss with direct regression of those angles,

$$\begin{aligned} \mathcal{L}_{\mathrm{FM}} &= \operatorname*{mean}_{k\in\mathcal{L}} \|\hat{v}_{\ell,k} - v_{\ell,k}\|^2 + \operatorname*{mean}_{k\in\mathcal{A}} \|\hat{v}_{\alpha,k} - v_{\alpha,k}\|^2 + \operatorname*{mean}_{k\in\mathcal{F}} \|\hat{v}_{\tau,k} - \Delta_{\tau,k}\|^2, \\ \mathcal{L}_1 &= \mathcal{L}_{\mathrm{FM}} + \operatorname*{mean}_{k\in\mathcal{R}} \left[(\hat{q}_{k,1} - \sin\tau_k)^2 + (\hat{q}_{k,2} - \cos\tau_k)^2 \right], \end{aligned} \tag{6}$$

where $\mathcal{L}$, $\mathcal{A}$ and $\mathcal{F}$ hold the generated bond lengths, bond angles and torsion angles, and $\mathcal{R}$ the rigid torsion angles. The second stage adds a loss on the distances between heavy atoms. The reason is the reconstruction. The natural extension reference frame (Parsons et al., 2005), or NeRF, constructs Cartesian coordinates from internal coordinates by placing each atom from the three before it. An error early in that walk carries every atom placed after it. $\mathcal{L}^{(1)}$ is an independent squared error per internal coordinate and never sees the structure they build together. The second stage therefore estimates the endpoint $\hat{x}_1 = x_t + (1-t)\,\hat{v}(x_t, t, G)$ from the interpolated state, where $\hat{v} = (\hat{v}_\ell, \hat{v}_\alpha, \hat{v}_\tau)$ stacks the three velocity heads, rebuilds the heavy-atom positions $\hat{p}$ from $\hat{x}_1$ by NeRF, and scores the distances there. Both sides are computed, $\hat{d}_{ij} = \|\hat{p}_i - \hat{p}_j\|$ on the rebuilt structure and $d^{\mathrm{ref}}_{ij} = \|p_i - p_j\|$ on the reference conformer. With $\mathcal{P} = \{(i,j) : i \neq j,\ d^{\mathrm{ref}}_{ij} < 8\,\text{Å}\}$ the heavy-atom pairs of one molecule, a cutoff Appendix C.4 reads off the references,

$$\mathcal{L}_{\mathrm{dist}} = \operatorname*{mean}_{\text{batch}} \operatorname*{mean}_{(i,j)\in\mathcal{P}} \left| \hat{d}_{ij} - d^{\mathrm{ref}}_{ij} \right|, \qquad \mathcal{L}_2 = \mathcal{L}_1 + \mathcal{L}_{\mathrm{dist}}. \tag{7}$$

3.4 Inference

Sampling. Each bond length and each bond angle is drawn from its own Gaussian and each torsion angle from the uniform circular prior, the priors of equation (5). We integrate the velocity field with twenty Euler steps. Every internal coordinate is updated at each step, and the torsion angles are then wrapped back to $[-\pi, \pi)$, since they live on the circle. The rigid torsion angles are overwritten by the regression head, as in training. The model regresses the conditional mean of the velocity, and integration is deterministic once x_0 is drawn, so the prior draw is the only randomness in a conformer. Renoising adds a second source. We restart at six times between $t = 0.25$ and $t = 0.75$. A restart draws a fresh prior sample x_0' and moves the state back from the current endpoint estimate $\hat{x}_1$ toward x_0', by the same $(1-t)$ the interpolation uses in training. The conformer found so far is kept and perturbed, and the state lands where the network saw states at this flow time during training. This costs six extra evaluations, 26 in place of 20. Equation (12) in Appendix C.5 gives the update. Every conformer is drawn independently and none is filtered out.

Closing the ring. The generated internal coordinates are rebuilt into Cartesian coordinates by NeRF. Reading only three atoms back, the walk ends as an open chain, the gap drawn in the inference panel of Figure 1. We continue it three atoms past the end of the ring, so the walk places a second copy of the three initial atoms, and solve for the internal coordinates that bring the copies onto the original Cartesian coordinates (Coutsias et al., 2004). Six is all the solver may change, and which six and why are worked out in Appendix F. A damped Gauss–Newton step does the solving, capped at forty iterations and regularized with weight 0.3 toward the values flow matching proposed.

Steric refinement. The all-atom structure is then rebuilt from the closed ring, and a short steric refinement pulls apart the close contacts that remain. It is post-processing with no learned part. Two hundred Adam steps on the Cartesian coordinates minimize three squared penalties. The first pushes apart any heavy-atom pair three or more bonds apart that sits within 2.3 Å and leaves a pair already further out alone. The second pulls every bond length and every bond angle back to the value the reconstructed conformer carries. The third holds each atom near where the reconstruction placed it. Every target is taken from that conformer rather than a force field or a reference. The one exception is the closing bond, held at the training set mean so the ring stays closed. Equation (14) in Appendix C gives the objective.

Table 1: Agreement with the reference ensembles. Each of the last four rows changes a single component of the full CYCLOFLOW and keeps the others. Bold is best and underline second best over CYCLOFLOW and the baselines.

		COV-F1 (%) ↑		MAT (Å) ↓		W_1 (Å) ↓		Clash
Method	**Params**	**Ring**	**All-atom**	**Ring**	**All-atom**	**Ring**	**All-atom**	**(%) ↓**
Reference oracle	–	$96.7_{\pm 0.1}$	$94.1_{\pm 0.2}$	$0.019_{\pm 0.000}$	$0.127_{\pm 0.001}$	$0.010_{\pm 0.001}$	$0.123_{\pm 0.002}$	$0.00_{\pm 0.00}$
ETKDGv3	–	$0.0_{\pm 0.0}$	$0.6_{\pm 0.0}$	$0.661_{\pm 0.001}$	$1.647_{\pm 0.001}$	$0.839_{\pm 0.001}$	$2.089_{\pm 0.008}$	$1.23_{\pm 0.04}$
+ MMFF94	–	$9.3_{\pm 0.0}$	$7.6_{\pm 0.1}$	$0.491_{\pm 0.001}$	$1.469_{\pm 0.002}$	$0.789_{\pm 0.002}$	$2.048_{\pm 0.005}$	$\mathbf{0.00}_{\pm 0.00}$
Torsional Diffusion	1.1M	$0.0_{\pm 0.0}$	$0.3_{\pm 0.1}$	$0.656_{\pm 0.001}$	$1.673_{\pm 0.014}$	$0.831_{\pm 0.000}$	$2.110_{\pm 0.014}$	$26.14_{\pm 7.14}$
RINGER	12M	$\underline{67.6}_{\pm 1.0}$	$\underline{64.8}_{\pm 0.2}$	$\underline{0.293}_{\pm 0.003}$	$\underline{0.810}_{\pm 0.005}$	$\underline{0.458}_{\pm 0.021}$	$\underline{1.264}_{\pm 0.009}$	$21.28_{\pm 1.37}$
MCF-S	13M	$27.5_{\pm 3.2}$	$31.4_{\pm 3.2}$	$0.368_{\pm 0.011}$	$1.180_{\pm 0.042}$	$0.681_{\pm 0.022}$	$1.842_{\pm 0.042}$	$26.90_{\pm 18.45}$
ET-Flow-*SO(3)*	9.1M	$35.7_{\pm 2.1}$	$31.8_{\pm 1.3}$	$0.350_{\pm 0.002}$	$1.166_{\pm 0.010}$	$0.675_{\pm 0.007}$	$1.856_{\pm 0.013}$	$5.47_{\pm 0.63}$
AGDiff	1.3M	$4.9_{\pm 0.3}$	$1.7_{\pm 0.1}$	$0.563_{\pm 0.004}$	$1.784_{\pm 0.009}$	$0.909_{\pm 0.005}$	$2.416_{\pm 0.019}$	$5.13_{\pm 0.29}$
$\text{AvgFlow}_{\text{DiT}}$	52M	$45.7_{\pm 1.1}$	$43.1_{\pm 1.3}$	$0.338_{\pm 0.006}$	$1.085_{\pm 0.016}$	$0.600_{\pm 0.035}$	$1.669_{\pm 0.051}$	$\underline{0.02}_{\pm 0.01}$
CYCLOFLOW	32M	$\mathbf{72.4}_{\pm 1.3}$	$\mathbf{75.2}_{\pm 0.3}$	$\mathbf{0.164}_{\pm 0.005}$	$\mathbf{0.555}_{\pm 0.007}$	$\mathbf{0.198}_{\pm 0.012}$	$\mathbf{0.841}_{\pm 0.006}$	$1.35_{\pm 0.39}$
First stage only	32M	$68.2_{\pm 1.2}$	$72.5_{\pm 0.4}$	$0.175_{\pm 0.001}$	$0.589_{\pm 0.005}$	$0.262_{\pm 0.012}$	$0.983_{\pm 0.032}$	$1.87_{\pm 0.31}$
No steric refinement	32M	$72.4_{\pm 1.3}$	$75.1_{\pm 0.3}$	$0.165_{\pm 0.005}$	$0.556_{\pm 0.008}$	$0.200_{\pm 0.013}$	$0.845_{\pm 0.007}$	$4.56_{\pm 0.79}$
No closure solve	32M	$31.3_{\pm 0.6}$	$73.8_{\pm 0.4}$	$0.229_{\pm 0.005}$	$0.617_{\pm 0.008}$	$0.312_{\pm 0.014}$	$0.898_{\pm 0.010}$	$1.68_{\pm 0.21}$
SLSQP closure	32M	$47.5_{\pm 0.5}$	$65.3_{\pm 0.5}$	$0.327_{\pm 0.005}$	$0.817_{\pm 0.010}$	$0.569_{\pm 0.010}$	$1.339_{\pm 0.017}$	$16.88_{\pm 0.35}$

4 RESULTS

Dataset and evaluation. We train and evaluate on CREMP (Grambow et al., 2024), 36,198 head-to-tail cyclic peptides of four to six residues. The peptides are built from 54 residue types covering L- and D-amino acids and their N-methylated variants. We follow the molecule-level partition of RINGER (Grambow et al., 2023), which holds out 1,000 cyclic peptides (533 tetra-, 347 penta-, 120 hexapeptides) and leaves 35,198, which we split into 31,679 training and 3,519 validation molecules. We keep the highest Boltzmann-weighted conformers of each molecule and use heavy atoms only, under the cap of Appendix C. For each test molecule we generate twice as many conformers as it has references. Every RMSD is minimized over the molecule's graph automorphisms. Every method is trained under the same three seeds and sampled once each. Numbers are means over the test set and those three runs, with W_1 a median over molecules instead, and the standard deviation covers training as well as sampling. Appendix G defines every metric used in this section and the next.

4.1 CYCLOFLOW IS CLOSEST TO THE REFERENCE ENSEMBLES

We compare each method to the reference ensemble with four metrics. Coverage (COV) and matching (MAT) follow RINGER, thresholds included, and both are built from nearest-neighbor distances. For each reference conformer we take the distance to the closest generated one, and for each generated conformer the distance to the closest reference. COV-F1 is the fraction of those distances below a threshold, 0.10 Å on the ring and 0.75 Å over all heavy atoms, and MAT is their mean. Since a nearest neighbor does not compare ensembles as distributions, we add the Wasserstein-1 distance (W_1) (Villani et al., 2009) between the generated conformers and the reference ones weighted by their Boltzmann weights. Unlike COV, W_1 charges for how much of the ensemble sits in each basin.

CYCLOFLOW is best on COV-F1, MAT and W_1, on the ring and over all heavy atoms (Table 1). On the ring, CYCLOFLOW leads RINGER by 4.8 points of COV-F1 and by more than twofold in W_1. A gap to the best attainable score remains, which the first row of the table measures. The conformer closest on the ring is not always the one closest over all heavy atoms, so Appendix I also measures all heavy atoms after superimposing on the ring alone, and Appendix J reports recall and precision separately by residue count. Physical validity is the remaining question. We count a conformer as clashing if any pair of heavy atoms at least three bonds apart lies closer than 2.0 Å, a cutoff no reference conformer meets. CYCLOFLOW clashes in 1.4% of its conformers, which is the price of generating in internal coordinates. ETKDGv3 + MMFF94 records 0.00% because minimizing MMFF94 removes such contacts outright through its van der Waals term. Appendix K gives W_1 and the clash rate at each residue count.

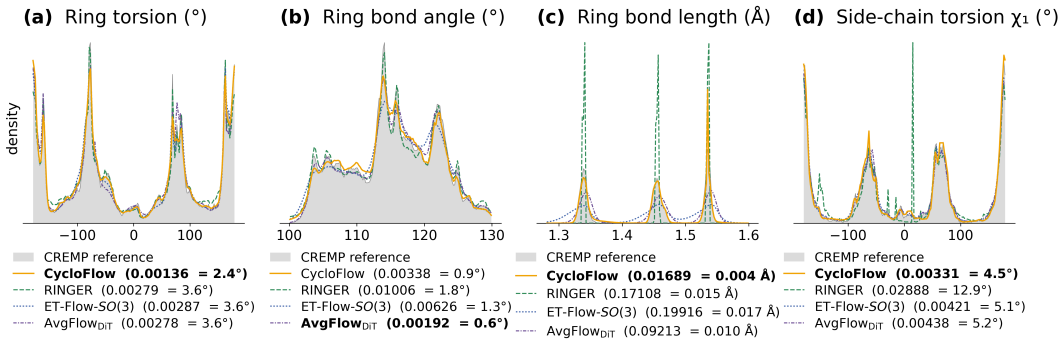

Figure 2: Distribution of each generated internal coordinate. (a) Ring torsion angle. (b) Ring bond angle. (c) Ring bond length. (d) Side-chain torsion angle χ_1. Parentheses give the Jensen–Shannon divergence to the reference and, after it, the width of the Gaussian blur that reproduces that divergence when applied to the reference coordinate. Bold is the lowest divergence in the panel.

4.2 What each component is worth

Each of the last four rows of Table 1 changes one component of CycloFlow and keeps the rest. Without the distance loss of the second training stage the model already leads every baseline on all six columns. Since the loss acts on the whole structure rather than on one coordinate at a time, adding it improves every column. The steric refinement acts on contacts alone. Removing it raises the clash rate from 1.4% to 4.6%, while COV, MAT and W_1 change within run-to-run variation. Appendix L splits that rate by contact type and sets it beside four xTB ensemble properties. The refinement clears the side-chain contacts, 4.5% to 0.5%, and leaves every property within its own spread, so it lowers the clash rate without moving the ensemble in energy.

The Gauss–Newton ring closure matters most on the ring. Removing the solve costs 41.1 points of ring COV-F1 but only 1.4 over all heavy atoms. MAT degrades by about the same distance on both, 0.065 Å on the ring and 0.062 Å overall, and the thresholds turn that into the gap in COV-F1. The solve brings the median ring to 0.090 Å of the nearest reference ring, from 0.264 before it runs. RINGER's SLSQP closure is the natural comparison. It adjusts ring atom positions, whereas our solve adjusts the six internal coordinates that closure determines. It shifts the ring torsion angles by a median $22.0°$ where ours shifts them by $7.7°$, and leaves the median ring at 0.280 Å. Paired per conformer, it moves a ring toward the reference by 0.012 Å where ours moves it by 0.053Å. Moving the ring atoms also moves the whole molecule with them, so the all-atom distance reaches 1.162Å where ours holds 0.708 and the clash rate 16.9%.

4.3 Each internal coordinate follows the reference distribution

Every metric so far scores whole conformers. Whether each internal coordinate is produced as a distribution at all is a separate question. We take each one on its own, pool it over the test set, and measure the Jensen–Shannon divergence (Lin, 1991) to the reference (Figure 2). A divergence has no units, so we invert it. Blurring one coordinate of the reference by a Gaussian of width σ produces a divergence, and the map is monotone, so each value in Figure 2 translates back into the blur that would produce it. Rebuilding the blurred coordinates turns the same σ into a displacement.

Read that way, CycloFlow reaches $2.4°$ on ring torsion angles versus $3.6°$ for the other three methods. On ring bond angle $\text{AvgFlow}_{\text{DiT}}$ leads at $0.6°$ against our 0.9. On χ_1 we reach $4.5°$ and RINGER 12.9. Bond lengths separate the methods furthest, 0.004 Å for CycloFlow against 0.015 for RINGER, and the shape of that failure is the point. RINGER's distribution is a spike on the per-bond-type training mean it restores rather than generates, so the coordinate carries no ensemble variation at all. These differences in individual coordinate distributions account for only part of the gap in conformer quality. The ring torsion angle difference corresponds to just 0.06 Å of structural displacement, compared with a ring MAT gap of 0.19 Å between the same methods. This suggests that how coordinate values combine within a ring reveals clearer differences between methods. Section 5.1 measures those combinations.

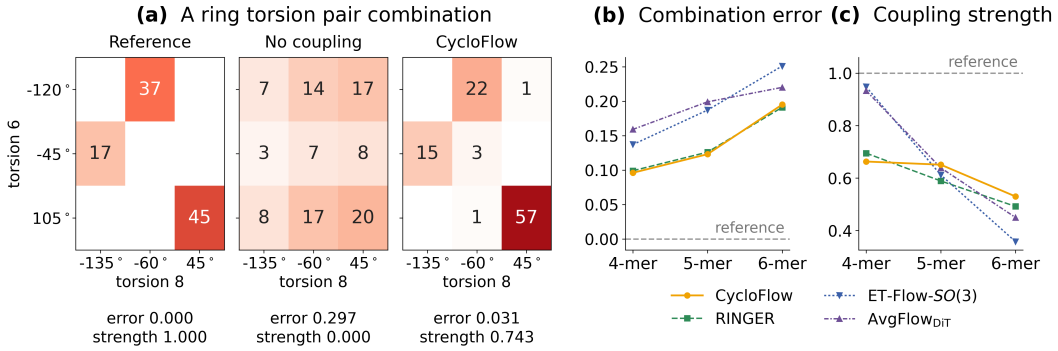

Figure 3: Ring torsion combinations. (a) A ring torsion pair of one test peptide. (b, c) Combination error and coupling strength of ring torsion pairs over the test molecules.

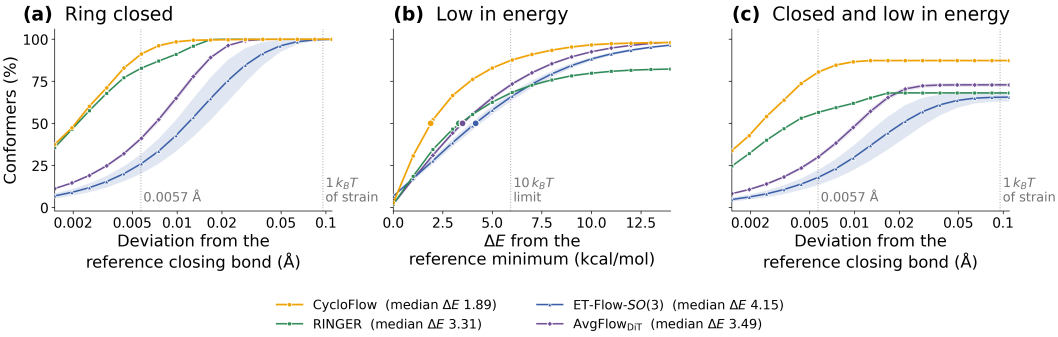

Figure 4: Ring closure and conformer energy. The percentage of conformers meeting a condition. (a) Deviation from the reference closing bond length, with a guide at the deviation that costs $1\,k_BT$ of MMFF94 strain (Appendix O). (b) Energy difference between a relaxed conformer and its molecule's reference minimum. (c) Both. Bands are the spread over runs.

5 Analysis

5.1 Coupling between ring torsions

Since the ring is closed, the ring torsions that are free to turn do not turn independently. Figure 3a takes the torsion 6 and torsion 8 pair of one test cyclic peptide. Each torsion settles into a few preferred wells, and only three of the nine combinations of those wells carry the reference ensemble, where pairing the two at random would spread that mass over all nine. We compare a generated ensemble against the reference two ways, over every pair of ring torsions that freely turn. The *combination error* compares the whole set of combinations of each torsion pair and reaches 0 when a model uses the reference's combinations in the reference's proportions. The *coupling strength* asks whether setting torsion 6 to one well determines where torsion 8 sits, and reaches 1 when the pair moves together as tightly as in the reference. CycloFlow and RINGER record almost the same combination error at every residue count (Figure 3b), and CycloFlow holds more of the coupling on 5-mers and 6-mers (Figure 3c). On 4-mers it holds less than any baseline while scoring the lowest error, since a pair can reach that mark and still sit in the wrong wells (Appendix M). Ring torsions couple to one another this way where side chains do not, so side chains are taken one torsion at a time in Appendix N.

5.2 Ring closure and conformer energy

We plot the share of conformers whose closing bond lands inside a widening tolerance of the reference (Figure 4a). A deviation under $1\,k_BT$ of strain is inside the molecule's own thermal motion, which puts that bar at 0.0955 Å. Every method stays inside it. Appendix O separates our two closure

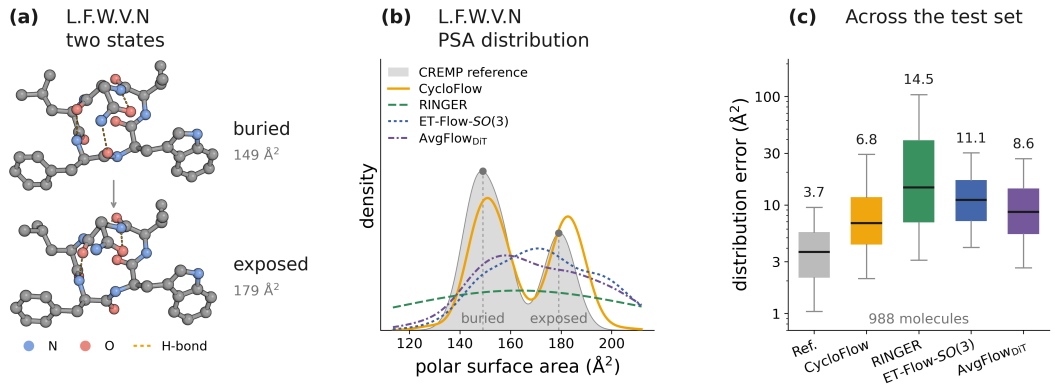

Figure 5: Polar surface area of the generated ensembles. (a) One reference molecule at the two modes of its ensemble, hydrogen bonds dashed. (b) Its polar surface over that ensemble. (c) Wasserstein distance to the reference distribution, over 988 test molecules.

components. Removing the steric refinement leaves the ring closed while removing the solve does not. At 0.0057 Å CycloFlow closes 90.9% of the time and RINGER 82.7%, while AvgFlow$_{\text{DiT}}$ and ET-Flow-*SO(3)* fall away. Our steric refinement targets the training-split mean bond length, as does RINGER's SLSQP closure. Whether the molecule would adopt the conformer that bond closes is a separate question. Panel (b) plots the same share against ΔE, a conformer's relaxed GFN2-xTB energy relative to its molecule's reference minimum. No reference conformer lies more than $10\,k_BT$ above it, which sets the limit (Appendix H). 87.3% of our conformers fall below that limit, and we lead at every tighter cutoff. 80.3% meet both criteria, the highest of the four.

5.3 Buried and exposed states in the ensemble

Membrane permeability depends on solvent-accessible polar surface and intramolecular hydrogen bonding, both of which vary across conformers. A single predicted structure captures only one of a peptide's buried or exposed states. Figure 5(a,b) shows a test molecule with two states differing by $30\,\text{Å}^2$ and an extra hydrogen bond in the buried state. Only CycloFlow separates them, and all three baselines fill the valley between. Panel (c) asks the same of the test molecules. Each conformer has a polar surface, so an ensemble has a distribution of them, and we measure the Wasserstein distance between the generated and the reference distribution, twenty conformers drawn from each so the sample sizes match. The floor of that measurement comes from splitting one reference ensemble in two, which needs forty reference conformers, and 988 of the 1,000 test molecules have them, so the panel covers those 988. CycloFlow reaches $6.8\,\text{Å}^2$ on a floor of $3.7\,\text{Å}^2$, the closest of the four. Reproducing a reference structure and reproducing the property it carries are not the same thing, and our ensembles do both.

6 Conclusion

We present CycloFlow, which generates cyclic peptide conformers in internal coordinates. A flow-matching model transports a bond length, a bond angle and a torsion angle at every heavy atom, and a Gauss–Newton solve then closes the ring on six of them. CycloFlow leads every generator we compare against on coverage, on matching and on the Wasserstein distance to the reference ensembles. The ablation shows that this solve, moving only those six from the values the flow proposed, keeps what the flow produced, whereas SLSQP closure over all ring atom positions loses a third of the ring coverage and leaves one conformer in six clashing. The flow learns the internal coordinates jointly, placing pairs of ring torsion angles level with the strongest baseline, and ahead of it on the larger rings. The generated ensembles also separate the buried and exposed states that membrane permeability depends on. Limitations remain. The largest rings are where the gap to the reference is widest, and side chains reach the right rotamers in the wrong proportions. Appendix P discusses the limitations in full.

REPRODUCIBILITY STATEMENT

Appendix C.1 gives the dataset, the split and the conformer cap. Appendix C.2 and Appendix C.3 give the input representation and the velocity field, Appendix C.4 the two training stages with their losses and schedule, and Appendix C.5 the sampler, the closure solve and the steric refinement, each with the constants it runs at. Appendix E gives the hyperparameter search that chose the architecture and Appendix D the hardware, the baseline configurations and the wall-clock cost of training and generation. Appendix G defines every metric in the paper and Appendix H the GFN2-xTB protocol behind every energy. Every number we report is an average over the same three training seeds, and the standard deviation beside it covers training as well as sampling. The source code, the evaluation scripts and the generated conformer ensembles the tables are read from are included in the supplementary material.

THE USE OF LARGE LANGUAGE MODELS

We used large language models in this work, and we set out here what they did. They drafted and edited prose from notes and results we supplied, suggested the arrangement of the manuscript and its appendix, and polished the text for readability. They searched for and summarized related work, which we then read and checked against the original sources. They wrote and edited the plotting scripts behind every figure. Reading the draft against the code and the data, they found errors, including stale numbers and claims the measurements did not support. The model and the training code are our own, and we used large language models to debug that code rather than to write it. The research question, the choice of baselines, the ablations, and the evaluation protocol are our own. We have reviewed all AI-assisted work, and we checked every number, figure, and claim in this paper against the code and the data. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *Journal of chemical theory and computation*, 15 (3):1652–1671, 2019.

Robert S Cahn, Christopher Ingold, and Vladimir Prelog. Specification of molecular chirality. *Angewandte Chemie International Edition in English*, 5(4):385–415, 1966.

Zhonglin Cao, Mario Geiger, Allan Dos Santos Costa, Danny Reidenbach, Karsten Kreis, Tomas Geffner, Franco Pellegrini, Guoqing Zhou, and Emine Kucukbenli. Efficient molecular conformer generation with SO(3)-averaged flow matching and reflow. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 6522–6542. PMLR, 13–19 Jul 2025. URL `https://proceedings.mlr.press/v267/cao25b.html`.

Evangelos A Coutsias, Chaok Seok, Matthew P Jacobson, and Ken A Dill. A kinematic view of loop closure. *Journal of computational chemistry*, 25(4):510–528, 2004.

Edward M Driggers, Stephen P Hale, Jinbo Lee, and Nicholas K Terrett. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nature Reviews Drug Discovery*, 7(7):608–624, 2008.

Colin A Grambow, Hayley Weir, Nathaniel L Diamant, Gabriele Scalia, Tommaso Biancalani, and Kangway V Chuang. Accurate and efficient structural ensemble generation of macrocyclic peptides using internal coordinate diffusion. *arXiv preprint arXiv:2305.19800*, 2023.

Colin A Grambow, Hayley Weir, Christian N Cunningham, Tommaso Biancalani, and Kangway V Chuang. Cremp: Conformer-rotamer ensembles of macrocyclic peptides for machine learning. *Scientific Data*, 11(1):859, 2024.

Thomas A Halgren. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *Journal of computational chemistry*, 17(5-6):490–519, 1996.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. Et-flow: Equivariant flow-matching for molecular conformer generation. *Advances in neural information processing systems*, 37:128798–128824, 2024.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in neural information processing systems*, 35:24240–24253, 2022.

Hajung Kim, Jueon Park, Junseok Choe, Seungheun Baek, Hyeon Hwang, and Jaewoo Kang. Graphcliff: Short-long range gating for modeling critical activity changes caused by subtle molecular differences. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '26, pp. 11260–11271, New York, NY, USA, 2026. Association for Computing Machinery. ISBN 9798400722592. doi: 10.1145/3770855.3818898. URL https://doi.org/10.1145/3770855.3818898.

Haitao Lin, Odin Zhang, Huifeng Zhao, Dejun Jiang, Lirong Wu, Zicheng Liu, Yufei Huang, and Stan Z. Li. Ppflow: target-aware peptide design with torsional flow matching. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Christopher A Lipinski, Franco Lombardo, Beryl W Dominy, and Paul J Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews*, 46(1-3):3–26, 2001.

Yunqing Liu, Yi Zhou, and Wenqi Fan. Geometric flow matching for molecular conformation generation via manifold decomposition. *arXiv preprint arXiv:2605.25577*, 2026.

Daniel S Nielsen, Nicholas E Shepherd, Weijun Xu, Andrew J Lucke, Martin J Stoermer, and David P Fairlie. Orally absorbed cyclic peptides. *Chemical reviews*, 117(12):8094–8128, 2017.

Jerod Parsons, J Bradley Holmes, J Maurice Rojas, Jerry Tsai, and Charlie EM Strauss. Practical conversion from torsion space to cartesian space for in silico protein synthesis. *Journal of computational chemistry*, 26(10):1063–1068, 2005.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4195–4205, October 2023a.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023b.

Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics*, 22(14): 7169–7192, 2020.

Vilya Research, :, Pascal Sturmfels, Milad Salem, Naozumi Hiranuma, Stephen Rettie, Xiaoliang Pan, Benjamin D. Sellers, Adam P. Moyer, Patrick J. Salveson, and Ivan Anishchanka. Vilya-1: An all-atom foundation model for macrocycle structure prediction and design, 2026. URL https://arxiv.org/abs/2607.09998.

Luca Schaufelberger, Aline Hartgers, and Kjell Jorner. Generating cyclic conformers with flow matching in cremer-pople coordinates. *arXiv preprint arXiv:2601.12859*, 2026.

Andre Brasil Vieira Wyzykowski, Fatemeh Fathi Niazi, and Alex Dickson. Agdiff: attention-enhanced diffusion for molecular geometry prediction. *Journal of Chemical Information and Modeling*, 65(4):1798–1811, 2025.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Alexander A Vinogradov, Yizhen Yin, and Hiroaki Suga. Macrocyclic peptides as drug candidates: recent progress and remaining challenges. *Journal of the American Chemical Society*, 141(10): 4167–4181, 2019.

Conan K Wang, Joakim E Swedberg, Peta J Harvey, Quentin Kaas, and David J Craik. Conformational flexibility is a determinant of permeability for cyclosporin. *The Journal of Physical Chemistry B*, 122(8):2261–2276, 2018.

Shuzhe Wang, Jagna Witek, Gregory A Landrum, and Sereina Riniker. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *Journal of chemical information and modeling*, 60(4):2044–2058, 2020.

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

Yuyang Wang, Ahmed A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Ángel Bautista. Swallowing the bitter pill: simplified scalable conformer generation. In *Proceedings of the 41st International Conference on Machine Learning*, ICML'24. JMLR.org, 2024.

Nicolas Wolf, Leif Seute, Vsevolod Viliuga, Simon Wagner, Jan Stühmer, and Frauke Gräter. Learning conformational ensembles of proteins based on backbone geometry. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 141926–141957. Curran Associates, Inc., 2025. doi: 10.52202/085713-4749. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/d02c860735807fd24a208525829b5ba0-Paper-Conference.pdf.

Guikun Xu, Xiaohan Yi, Ziqiao Meng, Peilin Zhao, and Yatao Bian. Energy-guided generative modeling for low-energy molecular structure discovery. *arXiv preprint arXiv:2512.22597*, 2025.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Xiangxin Zhou, Mingyu Li, Yi Xiao, Jiahan Li, Dongyu Xue, Zaixiang Zheng, Jianzhu Ma, and Quanquan Gu. Designing cyclic peptides via harmonic sde with atom-bond modeling. In *Proceedings of the 42nd International Conference on Machine Learning*, ICML'25. JMLR.org, 2025.

APPENDIX CONTENTS

A Background: What each generator produces and transports **14**

B Background: The measurements behind both choices **15**

C Implementation: Model, training and inference **16**

C.1 Data 16

C.2 Input representation 16

C.3 Velocity field 17

C.4 Training 17

C.5 Inference 18

D Implementation: Training and inference cost **19**

E Implementation: Hyperparameter search **20**

F Implementation: The six closure coordinates **20**

G Implementation: Metrics **21**

H Implementation: The xTB protocol **22**

I Additional results: All-atom scores under a ring-only fit **23**

J Additional results: Recall and precision by residue count **23**

K Additional results: W_1 and clashes by residue count **23**

L Additional results: xTB ensemble properties and Clashes **26**

M Additional analysis: Coupling between ring torsions **26**

N Additional analysis: Which rotamer a side chain occupies **26**

O Additional analysis: What counts as a closed ring **28**

P Limitations **29**

A BACKGROUND: WHAT EACH GENERATOR PRODUCES AND TRANSPORTS

This paper makes two choices, to generate internal coordinates and to transport them by flow matching. Each has been made before, never both together on a cyclic peptide. Table A1 places every method by what it produces and what transports it.

Cartesian coordinates, diffused or flowed. GeoDiff (Xu et al., 2022), MCF-S (Wang et al., 2024) and AGDiff (Vieira Wyzykowski et al., 2025) diffuse over atom positions, ET-Flow-*SO(3)* (Hassan et al., 2024), AvgFlow$_{\mathrm{DiT}}$ (Xu et al., 2025) and EnFlow (Xu et al., 2025) flow over them. Local geometry comes out with everything else and a closed ring is whatever the positions happen to give. All six were built for drug-like small molecules, whose rings are five- or six-membered and nearly rigid, so closure is not a modeling problem. EnFlow adds a learned energy that steers the sampler toward low-energy conformers, which is orthogonal to the choice made here and could sit on top of an internal-coordinate flow. Vilya-1 diffuses over atom positions as well and is the one Cartesian entry built for cyclic peptide ensembles (Research et al., 2026).

Internal coordinates, diffused. Torsional Diffusion takes local structure from an RDKit conformer and diffuses only the rotatable torsions, so a ring's own conformation is never sampled (Jing et al., 2022). RINGER diffuses ring bond angles as well and restores bond lengths from training set means (Grambow et al., 2023). It is the only generator we compare against that closes a ring, which it does by optimizing in Cartesian space afterwards.

Internal coordinates, flowed. PPFlow flows backbone torsion angles on the torus with lengths and angles held at ideal constants, for linear peptides, so there is no ring to close (Lin et al., 2024). GO-Flow flows all three internal coordinates but keeps a Cartesian state, so six rigid-body degrees of freedom ride along and have to be flowed too, and it has no closure stage (Liu et al., 2026). PuckerFlow works in Cremer–Pople coordinates, where a ring is closed by construction, for five- to eight-membered rings with lengths and angles fixed per bond type (Schaufelberger et al., 2026).

Built for a different task. BBFlow samples conformational ensembles of proteins over $SE(3)$ backbone frames, our task on a molecule with no ring (Wolf et al., 2025). CpSDE designs cyclic peptides, sequence and structure together, by a harmonic SDE over atoms and bonds (Zhou et al., 2025), and MuCO cyclizes a given linear peptide in three stages (Wang et al., 2026). Both make cyclic peptide structures, neither samples the ensemble of one that is given.

Table A1: *Velocity* is the space the model predicts in and *state* what the process carries from the prior to the sample, which decides whether a pose has to be generated and whether the ring can be left open. *gen.* marks a quantity the model produces, *fixed* one restored from ideal values or dataset means, and a dash one a Cartesian state leaves implicit. Appendix A reads the table.

				Predicted				
Method	Molecules	Velocity	State	length	angle	torsion	Transport	Ring conformation
ETKDGv3	small molecules, macrocycles	—	—		—		distance geometry	distance bounds
Torsional Diffusion	small molecules	internal	internal	fixed	fixed	gen.	diffusion	from RDKit
RINGER	cyclic peptides	internal	internal	fixed	gen.	gen.	diffusion	Cartesian SLSQP
PPFlow	linear peptides	internal	internal	fixed	fixed	gen.	flow matching	none
PuckerFlow	small molecules	Cremer–Pople	Cremer–Pople	fixed	fixed	gen.	flow matching	closed by construction
Vilya-1	cyclic peptides, ensembles	Cartesian	Cartesian		—		diffusion	implicit
GeoDiff	small molecules	Cartesian	Cartesian		—		diffusion	implicit
MCF-S	small molecules	Cartesian	Cartesian		—		diffusion	implicit
AGDiff	small molecules	Cartesian	Cartesian		—		diffusion	implicit
ET-Flow-*SO(3)*	small molecules	Cartesian	Cartesian		—		flow matching	implicit
AvgFlow$_{\mathrm{DiT}}$	small molecules	Cartesian	Cartesian		—		flow matching	implicit
EnFlow	small molecules	Cartesian	Cartesian		—		flow matching	implicit
GO-Flow	small molecules	internal	Cartesian	gen.	gen.	gen.	flow matching	not addressed
BBFlow	proteins, ensembles	SE(3) frames	SE(3) frames		—		flow matching	no ring
CpSDE	cyclic peptides, design	Cartesian	Cartesian		—		harmonic SDE	bond terms
MuCO	cyclic peptides, cyclization	SE(3) frames	SE(3) frames		—		flow matching	force field
CYCLOFLOW (ours)	cyclic peptides, ensembles	internal	internal	gen.	gen.	gen.	flow matching	six-coordinate solve

What is left empty is narrower than either axis and is what a cyclic peptide needs. A generator has to produce all three internal coordinates and close a ring too large for a puckering basis to span. GO-Flow produces all three and closes nothing, PuckerFlow closes by construction and fixes two of the three, PPFlow fixes two and has no ring at all, and every other internal coordinate entry fixes at least one. None of them generates a ring conformation and then closes it.

B Background: The measurements behind both choices

Every measurement below is a median over the 31,679 training molecules, so nothing here is read off the test set. Table A2 collects them.

What varies is torsional, and the other two still have to be right. Within one molecule's reference ensemble a ring bond length moves by 0.0018 Å and a ring bond angle by $0.63°$, against $8.61°$ for a ring torsion free to turn. Rebuilding two conformers of one molecule with lengths and angles held at the ensemble mean, so only torsion angles differ, reproduces the heavy atom RMSD between them to 100.6%, so torsion angles account for all of it, where the converse experiments recover 8.8% and 1.1%. Torsion angles are what a generator has to vary, and lengths and angles are what it has to get right instead. Section 4.3 finds that the methods restoring them from training means, or never representing them, are the ones that miss their distributions.

One torsion moves the whole ring. Rotating one ring torsion by $1°$ displaces the average heavy atom by 27.8 mÅ and the furthest by 100.0 mÅ. The same rotation on a side-chain torsion moves the average atom by 2.0 mÅ, a factor of 14.1 less, since a side chain has only its own branch downstream while a ring torsion has the whole cycle. A Cartesian generator has to learn that coupling from atom positions. In internal coordinates one torsion is one coordinate and the reconstruction propagates it, so there is nothing to learn.

The three move on different scales. Those three spreads differ by orders of magnitude, so one transport shared across them is scaled to whichever it was tuned for. Rescaling each coordinate before a shared schedule absorbs that. It does not absorb the periodicity of a torsion angle, whose value runs over the whole circle once molecules and wells are taken together, which is why flow matching lets each coordinate keep a prior and a path of its own.

Not every ring torsion is free. Planar amide bonds and proline hold 218,173 of the training set's 440,865 ring torsions rigid, and those move by $3.83°$, which is why their angles are regressed rather than carried by the velocity field.

Table A2: What varies within a CREMP ensemble, and what one torsion reaches. *Spread* is how far a ring coordinate moves inside one molecule's reference ensemble and *RMSD* how much of the distance between two of its conformers survives when only that coordinate differs. The free and rigid rows split the ring torsion angles by the model's own rotatable flag. On the right, how far one $1°$ rotation moves the average and the furthest heavy atom. Medians over the 31,679 training molecules.

Where an ensemble varies			**Reach of one $1°$ rotation**		
Coordinate	**Spread**	**RMSD**	**Torsion**	**Mean**	**Max**
Bond length	0.0018 Å	1.1 %	Ring	27.8 mÅ	100.0 mÅ
Bond angle	$0.63°$	8.8 %	Side chain	2.0 mÅ	–
Torsion	–	**100.6** %	ratio	$14.1\times$	
free	$8.61°$	–			
rigid	$3.83°$	–			
All three	–	100.0 %			

C Implementation: Model, training and inference

C.1 Data

CREMP stores a full CREST ensemble for every molecule and gives each conformer a GFN2-xTB free energy, which fixes its Boltzmann weight at 298 K. The median molecule holds 511, 856 and 1,006 unique conformers at four, five and six residues. We keep the $K = 40$ highest-weighted of each as the references, for training and for evaluation alike, which covers 92.4% of the Boltzmann mass and 6.2% of the distinct conformers. Figure A1 gives what other caps would cover, and returns fall off around forty. The training, valid and test splits are cut the same way.

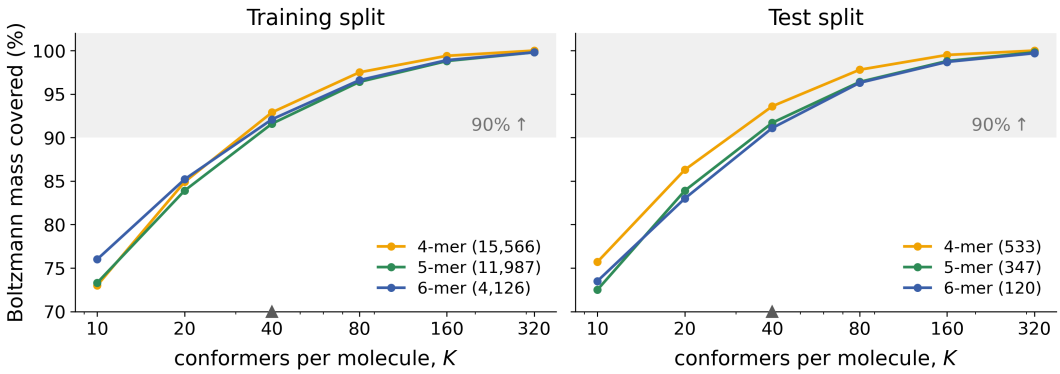

Figure A1: Boltzmann mass a cap of K conformers keeps, as a median over molecules. Each molecule's full CREST ensemble is sorted by weight and the top K are summed. The cap we use is $K = 40$, and the shading marks what a larger cap would add.

C.2 Input representation

Molecular features. The feature vector f_k is five-dimensional, holding the atom's degree, aromaticity, ring membership, hydrogen count and CIP code, and the flags r_k and s_k add one each. Each bond feature e_{ij} is four-dimensional, marking single, multiple, aromatic and in-ring bonds.

Which torsion angles are rigid. The flag r_k is set from the bond the torsion at atom k is measured on, before any conformer is seen, so it is a property of the graph. That bond is rigid when it is aromatic, double or triple, when it joins a carbonyl carbon to a nitrogen or an oxygen, which is the planar amide or ester, or when it lies in a ring of six atoms or fewer other than the macrocycle, which is what fuses a proline onto the backbone. Every other bond is rotatable. Half of the ring torsion angles come out rigid, 218,173 of the training split's 440,865, and those move by $3.83°$ within an ensemble against $8.61°$ for the ones free to turn.

Why the closure flag is hidden. The flag s_k of equation (2) keeps its slot in the node feature and in the graph encoder, and we hold both at zero, so the network never sees which six coordinates the closure solve will own. A network that sees the flag keys on it and ties itself to the one closure window it trained with, which would fix that window at inference. Zeroing the flag rather than dropping it leaves the input width, and with it the checkpoint, unchanged.

Where the ring closes. The atom order puts the closing bond, the one the reconstruction never places, on a C–C bond in every molecule of the dataset, joining a carbonyl carbon to the next α carbon. The closure residual is written where that bond falls and the six solved coordinates are taken next to it, so what matters is whether the coordinates there are free to move. About a C–C closing bond the torsion is a backbone ψ, which turns freely, and the bond angles at its ends are ordinary. The alternative leaves the amide C–N to close, where the torsion is ω, one of the rigid angles the model regresses rather than transports, and the angles at its nitrogen are held by the same planarity. The solve would then have to bend a planar group at the one place it cannot avoid touching.

Graph encoder. Two GraphCliff layers (Kim et al., 2026). The per-atom input is f_k, the two flags and a one-hot composite type crossing the atomic number with the aromatic and in-ring flags and with ring against side chain, 329 dimensions in all, and the output is d_{model} wide. It reads no internal coordinate, so $\text{GE}(G)$ is computed once per molecule and reused at every step and by every conformer.

Atom embedding. Equation (2) sums three vectors of width $d_{\text{model}} = 384$,

$$h_k^{(0)} = \underbrace{\text{Emb}(Z_k)}_{40 \to 384} + \underbrace{W_{\text{node}}\left[f_k,\ r_k,\ s_k,\ \ell_k^{(t)},\ \alpha_k^{(t)},\ \sin\tau_k^{(t)},\ \cos\tau_k^{(t)}\right]}_{11 \to 384} + \underbrace{W_{\text{ge}}\,\text{GE}(G)_k}_{384 \to 384}. \quad (8)$$

Summing rather than concatenating keeps the width fixed however many features are added, and only the middle term changes along the trajectory.

C.3 Velocity field

An atom-level Transformer with hidden width $d_{\text{model}} = 384$, eight blocks and eight attention heads, 32.07 M parameters and a 384 MB checkpoint.

Time conditioning. The modulation MLP of equation (3) ends in one linear layer that emits all six parameters together, and that layer starts at zero. Both gates g_1^ℓ, g_2^ℓ are therefore zero at the start, so each block begins as an identity map and learns its conditioning gradually, as in DiT (Peebles & Xie, 2023b).

Transformer block. The per-head bias of equation (4) is read from the graph. With δ_{ij} the shortest-path distance between two atoms, capped at $\delta_{\max} = 16$ by a bucket that holds every pair beyond it, Emb_δ an embedding table over those buckets, and e_{ij} the bond feature,

$$a_{ij}^{(m)} = \frac{\left(Q^{(m)}h_i\right)^\top \left(K^{(m)}h_j\right)}{\sqrt{d_h}} + \beta_{ij}^{(m)}, \qquad \beta_{ij}^{(m)} = \text{MLP}\left([\text{Emb}_\delta(\delta_{ij}),\, e_{ij}]\right)_m, \quad (9)$$

one bias per head m, added to the logit before the softmax.

Output heads. Four heads read the final hidden state, each a two-layer MLP of width 384. Three emit one number apiece, the velocity of the bond length, of the bond angle and of the torsion angle. The fourth emits the two components of $\hat{q}_k$ at every rigid torsion angle. Nothing else is read off the network.

C.4 Training

Priors and weights. The torsion angle prior of equation (5) is uniform on the circle and has nothing to set. The bond angle and bond length priors are $(\mu_\alpha, \sigma_\alpha) = (1.95, 0.20)$ rad and $(\mu_\ell, \sigma_\ell) = (1.45, 0.15)$ Å, generic chemistry values rather than statistics of the training split. Every term of equation (6) carries unit weight.

Stage 1. Trained from scratch on $\mathcal{L}^{(1)}$ of equation (6) at a learning rate of 2×10^{-4}. Flow time is drawn uniformly on $[0, 1]$ for every example, and an example is one reference conformer, so a molecule contributes as many examples per epoch as it has kept conformers.

Stage 2. Warm-starts from the first and adds the pair-distance loss of equation (7) at a tenth of the learning rate. That loss is taken over Cartesian coordinates, so the internal ones are turned into them by the natural extension reference frame (Parsons et al., 2005). Writing A, B, C for the three atoms already placed and ℓ, α, τ for the new atom's own coordinates, with $\hat{b} = (C - B)/\|C - B\|$, $\hat{n}$ the unit normal of the plane ABC and $\hat{m} = \hat{n} \times \hat{b}$, that atom goes at

$$D = C + \ell\left(-\cos\alpha\,\hat{b} + \sin\alpha\cos\tau\,\hat{m} + \sin\alpha\sin\tau\,\hat{n}\right). \quad (10)$$

Three seed atoms start the walk and every later atom follows in one pass, so the structure is a composition of equation (10) and differentiable in the coordinates the flow produced. That is what lets the distance loss reach the velocity field, and the closure residual of equation (15) is written on the same reconstruction. The term earns its own stage. Trained from scratch alongside the flow-matching loss it was worse than that loss alone, and applied to a converged model at the lower learning rate it improved every column of Table 1, trading a little recall for a larger gain in precision.

The distance cutoff. That loss is taken over heavy-atom pairs closer than $8\,\text{Å}$ in the reference conformer. The cutoff is wide enough to cover most of a molecule this size and still leave the loss local, holding 89.8%, 84.3% and 82.3% of the heavy-atom pairs at four, five and six residues.

Schedule. Both stages run at an effective batch of 2,048 with a constant learning rate, a fresh optimizer at the second stage and an early stop on validation loss. The second stage has patience 10 and takes 67, 27 and 70 epochs over the three seeds, whose best checkpoints fall at epochs 56, 16 and 59 and reach validation losses of 0.8286, 0.8396 and 0.8261. The second-stage loss adds the all-atom distance term, so it is not on the same scale as the first stage's. The seeds differ in nothing else.

C.5 Inference

Sampling. Twenty Euler steps at unit velocity gain, so $\Delta t = 1/20$ and the state moves by

$$x_{t+\Delta t} = x_t + \Delta t\, v_\theta(x_t, t), \tag{11}$$

with the torsion rows wrapped after each update. At the six times $t \in \{0.25, 0.35, 0.45, 0.55, 0.65, 0.75\}$ we renoise. The model's endpoint estimate is $\hat{x}_1 = x_t + (1-t)\, v_\theta(x_t, t)$, and the state is replaced by the training interpolant at the same t between that estimate and a fresh prior draw x_0',

$$x_t \leftarrow \hat{x}_1 + (1-t)\,\mathrm{wrap}(x_0' - \hat{x}_1)\,. \tag{12}$$

All three coordinates are redrawn, so the state is one the model was trained to read rather than half-integrated, and the restarts take the network evaluations from 20 to 26. The rigid torsion angles are integrated with the rest and then overwritten by $\hat{q}_k$ once the last step is taken.

Closing the ring. Write u for the six closure internal coordinates, $r(u)$ for the residual of equation (15) and J for its Jacobian. With u_0 the values the flow proposed, each iteration solves

$$\left(J^\top J + (w^2 + \lambda) I\right) \Delta u = J^\top r + w^2\,(u - u_0)\,, \qquad u \leftarrow u - \Delta u, \tag{13}$$

with the torsion rows of $u - u_0$ wrapped to the circle. The closure condition has several solutions and the residual alone prefers none of them, so the proximity term, carrying $w = 0.3$, picks the one nearest the flow's proposal. The damping λ starts at 10^{-3}, halves after a step that lowers the residual and quadruples after one that does not, and a step that raises it is rejected. The solve stops at an RMS residual of 10^{-3} Å or after 40 iterations.

Steric refinement. Write $d_{ij} = \|x_i - x_j\|$ for the distance between two atoms, $\hat{d}_{ij}$ for its target, and $\mathcal{E}_1, \mathcal{E}_2$ and $\mathcal{E}_{\geq 3}$ for the pairs at graph distance $\delta_{ij} = 1$, $\delta_{ij} = 2$ and $\delta_{ij} \geq 3$. The first two are the bonded pairs and the two ends of a bond angle, so holding them at their targets holds the local geometry. The third are the pairs free to collide. With $\tilde{x}$ the reconstructed coordinates, the steric refinement of Section 3.4 minimizes

$$E(x) = w_{\text{rep}} \operatorname*{mean}_{(i,j)\in\mathcal{E}_{\geq 3}} \mathrm{relu}(c - d_{ij})^2 + \sum_{\mathcal{S}\in\{\mathcal{E}_1,\mathcal{E}_2\}} \operatorname*{mean}_{(i,j)\in\mathcal{S}} w_{ij} \left(d_{ij} - \hat{d}_{ij}\right)^2 + w_{\text{res}} \operatorname*{mean}_k \|x_k - \tilde{x}_k\|^2. \tag{14}$$

Two hundred steps of Adam at learning rate 0.02, with a flat-bottomed repulsion cutoff $c = 2.3\,\text{Å}$ and weights $w_{\text{rep}} = 20$, $w_{\text{res}} = 0.1$, and $w_{ij} = 10$ on $\mathcal{E}_1$ and 5 on $\mathcal{E}_2$. The closing bond is the one exception. Its target is the training-split mean $1.5348\,\text{Å}$ rather than the generated distance and its weight is $1{,}000$, while its two $\mathcal{E}_2$ partners keep generated targets, since constant ones would fix the closing bond angle as well and spike its distribution. That one bond is therefore held rather than

generated, and its spread within a molecule falls to 0.00008 Å against the references' 0.00147. The other ring bonds keep theirs, 0.00275 Å against 0.00181, where RINGER's restoration from training means leaves every ring bond at 0.00000. Every other target is read off the generated conformer, so the step needs no reference.

D Implementation: Training and inference cost

Tables A3 and A4 give what it costs to train each method and to generate the test ensembles. Most baselines train on A100s in their own papers, so we train CycloFlow and every baseline on A100 80GB, except AvgFlow$_{\mathrm{DiT}}$, whose long multi-GPU run goes on RTX PRO 6000. RINGER keeps the CREMP configuration it comes with. The others have none for cyclic peptides and follow their GEOM-Drugs one, whose molecules are the closest in size and flexibility, with three departures. AGDiff runs 1M of its 2M iterations because it is already converged. AvgFlow$_{\mathrm{DiT}}$ adds chirality to its atom features because the original code drop the chirality feature then the scores without chirality are much worse.

Training. An epoch means different things across these methods. Torsional Diffusion, ET-Flow-*SO(3)* and AvgFlow$_{\mathrm{DiT}}$ draw one conformer of each training molecule, RINGER and CycloFlow pass over every kept conformer, and MCF-S and AGDiff run for a set number of optimizer steps. The column that compares is therefore conformer passes, optimizer steps times effective batch over the 1.26M training conformers. Three caveats belong with the GPU-hours, which are wall-clock times the cards held.

Table A3: Training budget per seed, averaged over the three seeds. VRAM: peak per GPU.

Method	Params	GPU	Eff. batch	Epochs	Conf. passes	Opt. steps	GPU-h	VRAM (GiB)
ETKDGv3	–	–	–	–	–	–	–	–
+MMFF94	–	–	–	–	–	–	–	–
Torsional Diffusion	1.05M	A100	32	250	6.3	247.5k	20.5	24.8
RINGER	12.0M	A100	8,192	1,000	1,000	155.0k	96.0	46.7
MCF-S	12.9M	A100	32	19.0	19.0	750.0k	53.7	2.3
ET-Flow-*SO(3)*	9.06M	8×A100	384	200	5.0	16.6k	20.2	78.9
AGDiff	1.34M	A100	32	25.4	25.4	1,000k	74.7	16.9
AvgFlow$_{\mathrm{DiT}}$	52.4M	8×RTX PRO 6000	512	761	19.1	47.2k	243.3	48.8
CycloFlow								
First stage	32.1M	A100	2,048	329.0	328.9	202.7k	48.1	61.4
Second stage	32.1M	A100	2,048	54.7	54.7	33.7k	13.0	61.4

Inference. Every method generates the same 79,568 conformers, timed from the first network call to the last written one. Three rows need reading with care. ETKDGv3 runs on CPUs without time deadline to make all conformers. RINGER's time covers its official reconstruction CLI, diffusion sampling on the GPU (2.7 min), ring closure on 8 CPU cores (24 min) and side-chain rebuilding (2.4 min). Our own time includes the steric refinement.

Table A4: End-to-end inference cost over the 1,000 test peptides, averaged over the three seeds. VRAM: peak per GPU. The two ETKDG rows are one seed, re-timed on the same Xeon the other CPU rows use, so the comparison is not decided by core count.

Method	GPU	CPU	VRAM (GiB)	Generation	Relaxation	Per molecule
ETKDGv3	–	36 procs, Xeon Gold 5220R	–	0.27 h	–	0.96 s
+MMFF94	–	36 procs, Xeon Gold 5220R	–	0.27 h	0.21 h	1.73 s
Torsional Diffusion	A100	EPYC 7713	51.5	2.90 h	–	10.4 s
RINGER	A100	Xeon Gold 5220R, 8 cores	26.9	0.50 h	–	1.81 s
MCF-S	A100	EPYC 7713	3.0	6.47 h	–	23.3 s
ET-Flow-*SO(3)*	A100	Xeon Gold 5220R	61.0	3.34 h	–	12.0 s
AGDiff	A100	EPYC 7713	2.8	86.1 h	–	310 s
AvgFlow$_{\mathrm{DiT}}$	RTX PRO 6000	2×EPYC 9655	33.6	3.00 h	–	10.8 s
CycloFlow	A100	Xeon Gold 5220R	2.4	0.32 h	0.40 h	2.58 s

Table A5: Hyperparameter grid. The four terms above the rule were searched over 16 configurations and compared by validation ring COV-F1; selected values in bold. The terms below were held fixed. Each configuration was trained from scratch with early stopping on validation loss, at patience 30 in the first stage and patience 10 in the second.

Term	Values searched
Model width	320, **384**
Attention heads	**8**, 16
Pair-feature width	**6**, 8
GraphCliff embedding depth	**2**, 3
Held fixed	
Transformer blocks	8
Effective batch size	2,048
Learning rate	
- First stage	2×10^{-4}
- Second stage	2×10^{-5}
Bond-angle loss weight	1.0
Bond-length loss weight	1.0

E Implementation: Hyperparameter search

We tune CycloFlow over four architecture terms with two values each, 16 configurations in all (Table A5), model width, attention heads, pair-feature width and GraphCliff embedding depth. Every configuration trains from scratch with early stopping on validation loss at a patience of 30 epochs, and validation ring COV-F1 over seeds 10, 20 and 30 selects between them. The selected configuration reaches 72.63 ± 0.31 over the 3,519 validation macrocycles. No test molecule enters the selection. We hold the number of blocks, the effective batch, the learning rate and the bond-angle and bond-length loss weights fixed, and we do not tune the baselines, each keeping the configuration named in Appendix D.

F Implementation: The six closure coordinates

Write P_{seed} for the three atoms that start the reconstruction of Section 3.4 and P_{end} for the same three reached the long way around the ring. With u the six closure internal coordinates and each triple stacked as a 3×3 matrix, the solve is

$$r(u) = \mathrm{vec}\left(P_{\mathrm{end}}(u) - P_{\mathrm{seed}}\right), \qquad u^{\star} = \arg\min_{u} \|r(u)\|^2. \tag{15}$$

Bringing one rigid triple onto another is six conditions, three of position and three of orientation (Figure A2A, B). The residual is written over three atoms and so has nine Cartesian components, but those three are rigid with respect to one another by construction. Their three pairwise distances survive any rigid motion (Figure A2C), so three of the nine are redundant. Six independent conditions remain, the closure Jacobian has rank six, and six ring coordinates are determined once the rest are chosen.

The six cannot always be six torsion angles. The flexibility is torsional, so the solver takes torsion angles first, but a small ring does not have six rotatable ones. An amide C–N is planar and a proline fuses its own five-membered ring onto the backbone N–Cα bond, locking one more torsion apiece, which leaves the number of rotatable ring bonds R at 6 for a 4-mer without proline, 5 with one and 4 with two, rising to 8 for a 5-mer and 9 or more for a 6-mer.

A solver that took torsion angles until it had six would take all six of a 4-mer, leaving the flow no ring torsion angle to vary and the generator nothing to generate. We therefore reserve three for the flow, let the solver take $\min(6, R-3)$ of the rest, and top the set up to six with ring bond angles at the closing bond. A 4-mer without proline gives the solver three torsion angles and three bond

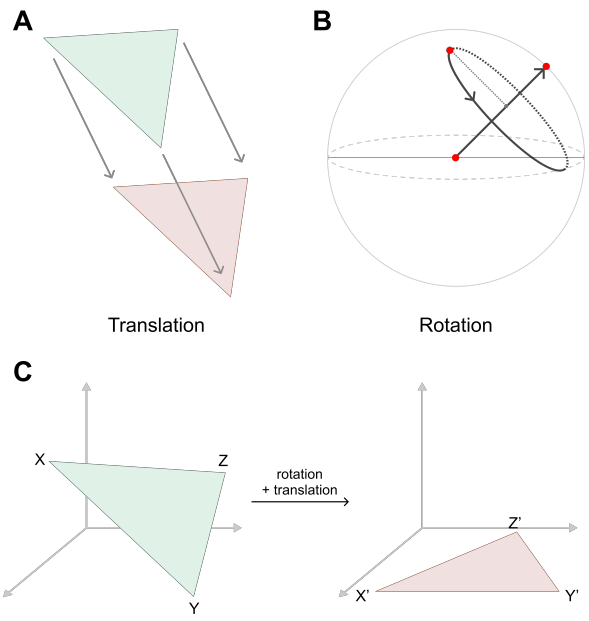

Figure A2: Six degrees of freedom of a rigid frame in three dimensions. (A) Translation moves the frame without changing it, which is three degrees of freedom. (B) Rotation is three more. Fixing one point leaves the second on a sphere around it, two angular degrees of freedom, and fixing the axis through the two leaves the third on a circle about that axis, one more. (C) A rigid motion combines the two and leaves the three pairwise distances of the frame unchanged, which is what makes three of the nine closure conditions redundant.

angles, a 6-mer gives it six torsion angles and none. Only 117 of the 1,000 test molecules reach six torsion angles, all of them 6-mers.

G Implementation: Metrics

Write $R = \{r_1, \dots, r_M\}$ for a molecule's reference ensemble, $w_1, \dots, w_M$ for the Boltzmann weights of Appendix C.1 normalized to sum to one, and $G = \{g_1, \dots, g_N\}$ for the generated ensemble. All distances are

$$d(x, y) = \min_{\pi \in \mathrm{Aut}(G_{\mathrm{mol}})} \min_{T \in SE(3)} \sqrt{\tfrac{1}{|S|} \textstyle\sum_{k \in S} \| T x_k - y_{\pi(k)} \|^2}, \tag{16}$$

the heavy-atom RMSD after optimal superposition, minimized over the automorphisms of the molecular graph so that a ring flip or a pair of equivalent methyls is not charged as a different conformer. The atom set S is the ring for the ring column and every heavy atom for the all-atom column.

Coverage and matching. With a threshold δ,

$$\begin{aligned} \text{COV-R} &= \tfrac{1}{M} \left| \{\, i : \min_j d(r_i, g_j) < \delta \,\} \right|, \quad \text{MAT-R} = \tfrac{1}{M} \textstyle\sum_i \min_j d(r_i, g_j), \\ \text{COV-P} &= \tfrac{1}{N} \left| \{\, j : \min_i d(r_i, g_j) < \delta \,\} \right|, \quad \text{MAT-P} = \tfrac{1}{N} \textstyle\sum_j \min_i d(r_i, g_j). \end{aligned} \tag{17}$$

COV-F1 is the harmonic mean of the two coverages and MAT the arithmetic mean of the two distances. Following RINGER (Grambow et al., 2023), $\delta = 0.10$ Å on the ring and 0.75 Å over all heavy atoms.

Wasserstein distance. With cost $C_{ij} = d(r_i, g_j)$,

$$W_1 = \min_{P \geq 0} \textstyle\sum_{ij} P_{ij} C_{ij} \quad \text{subject to} \quad \textstyle\sum_j P_{ij} = w_i, \;\; \textstyle\sum_i P_{ij} = \tfrac{1}{N}, \tag{18}$$

solved exactly as a linear program per molecule. Each molecule gives one W_1 and the table reports the median over molecules, where COV and MAT are means. Mass sitting far from where the reference wants it is charged, and so is mass piled where the reference has little, neither of which a nearest-neighbor statistic sees.

Clashes. A conformer clashes if any pair of heavy atoms separated by three or more bonds in the molecular graph is closer than 2.0 Å. The rate is the fraction of generated conformers with at least one such pair.

Distributions of one coordinate. Each internal coordinate is pooled over the test set into a fixed histogram and compared by the Jensen–Shannon divergence (Lin, 1991), $\mathrm{JS}(p, q) = \frac{1}{2}\mathrm{KL}(p\|m) + \frac{1}{2}\mathrm{KL}(q\|m)$ with $m = \frac{1}{2}(p + q)$, symmetric so that a missed mode and an invented one cost alike.

Coupling. For one molecule, every ring torsion angle whose reference circular standard deviation exceeds 20° is assigned to a well taken from that molecule's own reference, and each pair of such angles gives a contingency table. Writing p_{ij} for the reference joint over two angles' wells, q_{ij} for the generated joint and $p_i p_j$ for the product of the reference marginals, the *combination error* is $\mathrm{JS}(q, p)$ and the *reference coupling* is $\mathrm{JS}(p_i p_j, p)$, which is how much joint structure there is to miss. The *coupling strength* is the mutual information of the generated table over that of the reference, and it reaches 1 when the generated pair carries as much dependence as the reference pair. It is zero for a generator whose two angles are independent, whichever wells they occupy, and a pair can carry the reference's dependence while occupying the wrong wells. Twenty samples give a positive mutual information even when two angles are independent, so each side has its own floor subtracted, measured by shuffling one angle within its own ensemble. Both sides are subsampled to 20 conformers, since a divergence between histograms moves with the number of samples that built them.

Polar surface and hydrogen bonds. The three-dimensional polar surface area of a conformer is the Shrake–Rupley solvent-accessible surface of its nitrogens, oxygens and the hydrogens bonded to them, with Bondi radii (1.20, 1.70, 1.55, 1.52 and 1.80 Å for H, C, N, O and S), a 1.4 Å probe and 192 test directions per atom placed on a golden spiral. A topological descriptor would give one number per molecule, where this varies between conformers, which is the point of measuring it. An intramolecular hydrogen bond is counted when an N–H donor and an N or O acceptor more than two bonds apart have $\mathrm{H}\cdots\mathrm{A}$ below 2.5 Å and a donor angle above 120°. Both read the GFN2-xTB relaxed structures of Appendix H, which is the only place the hydrogens are placed by anything better than geometry.

H Implementation: The xTB protocol

Every energy, dipole μ and HOMO–LUMO gap $\Delta\varepsilon$ in this paper is read off a structure relaxed with GFN2-xTB under the protocol below, and so is the polar surface of Section 5.3. Two choices decide the numbers.

The solvent model has to be CREMP's. That the references were optimized under ALPB chloroform is checkable rather than assumed. For `A.A.A.A` conformer 0 the energy stored in CREMP is $-63.97615198\,E_h$, `--alpb chcl3` returns $-63.97615111\,E_h$, and the gas phase returns $-63.94070597\,E_h$, 22 kcal/mol away. Relaxing on any other surface moves a conformer off the minimum the reference sits on, and it moves the margins with it. Our lead over RINGER on $E_{\min}$ widens from $3.9\times$ in the gas phase to $9\times$ under ALPB.

Conformer counts are equalized before anything is averaged, so the comparison is not decided by how many samples an arm holds. The reference is the dataset's top-K by Boltzmann weight and each method contributes its first $2K$. Each entry of Table A12 is the median over molecules of $|\text{generated} - \text{reference}|$, Boltzmann-averaged at 298.15 K.

The same run calibrates the energy limit of Section 5.2. Not one of the $39{,}784$ reference conformers lies more than $10\,k_BT$ (5.9 kcal/mol) above its own molecule's minimum, and 92% lie within $5\,k_BT$, so the limit is read off the references rather than chosen.

Table A6: All-atom scores under a ring-only fit. The superposition is fitted over the ring alone and the RMSD is then measured over every heavy atom, at a threshold of 0.75 Å, so a fit cannot be pulled off the ring to hide a misplaced side chain. Bold is best per column, underline second best.

	COV-F1 (%) ↑				MAT (Å) ↓			
Method	**All**	**4-mer**	**5-mer**	**6-mer**	**All**	**4-mer**	**5-mer**	**6-mer**
RINGER	$\underline{62.7}_{\pm 0.2}$	$\underline{70.6}_{\pm 0.5}$	$\underline{57.9}_{\pm 1.1}$	$\underline{40.6}_{\pm 1.5}$	$\underline{0.939}_{\pm 0.007}$	$\underline{0.765}_{\pm 0.019}$	$\underline{1.029}_{\pm 0.010}$	$\underline{1.454}_{\pm 0.025}$
ET-Flow-*SO(3)*	$27.8_{\pm 1.2}$	$40.9_{\pm 1.7}$	$15.1_{\pm 0.7}$	$2.4_{\pm 0.1}$	$1.344_{\pm 0.013}$	$1.059_{\pm 0.020}$	$1.506_{\pm 0.006}$	$2.142_{\pm 0.008}$
AvgFlow$_{\text{DiT}}$	$39.9_{\pm 1.3}$	$50.8_{\pm 1.5}$	$31.6_{\pm 1.4}$	$10.5_{\pm 1.4}$	$1.262_{\pm 0.020}$	$1.045_{\pm 0.034}$	$1.351_{\pm 0.002}$	$1.966_{\pm 0.028}$
CycloFlow	$\mathbf{73.0}_{\pm 0.5}$	$\mathbf{78.4}_{\pm 0.7}$	$\mathbf{72.5}_{\pm 0.6}$	$\mathbf{50.0}_{\pm 2.1}$	$\mathbf{0.641}_{\pm 0.009}$	$\mathbf{0.537}_{\pm 0.015}$	$\mathbf{0.620}_{\pm 0.007}$	$\mathbf{1.164}_{\pm 0.044}$

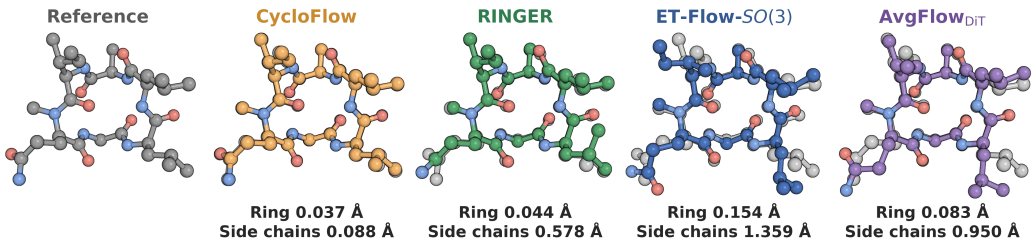

Figure A3: What the all-atom score hides. A 6-mer test peptide, reference in gray, with the closest generated conformer of each method superimposed on the ring alone. Every method places the ring to within 0.16 Å, where the side-chain distances span an order of magnitude, 0.088 Å for CycloFlow against 1.359 for ET-Flow-*SO(3)*.

I Additional results: all-atom scores under a ring-only fit

A superposition fitted over all heavy atoms can be pulled away from the ring to compensate for misplaced side chains, which flatters a conformer whose ring is right and whose substituents point the wrong way. Superimposing on the ring alone and measuring over all heavy atoms afterwards removes the compensation (Table A6, Figure A3).

Scored that way over all 1,000 test molecules, CycloFlow leads at every residue count under this fit, and by more than under a free one.

J Additional results: recall and precision by residue count

Table 1 pools residue counts and combines the two directions. The tables below open both up, at the same thresholds. Either direction can be raised on its own, since a model that generates indiscriminately covers more of the reference and one that generates a single safe conformer keeps its precision high, which is why the main table combines them.

K Additional results: W_1 and clashes by residue count

Appendix J splits coverage and matching by residue count. Table A11 does the same for the two quantities Table 1 leaves pooled. W_1 splits the way coverage and matching do. CycloFlow is best at every size, on the ring and over all heavy atoms, and the margin widens with the ring, from 0.158 against RINGER's 0.299 at four residues to 0.522 against 0.868 at six. Weighting the reference by its own Boltzmann populations reorders nobody. Clashes do not follow that pattern. Ours are worst on the smallest rings, 2.37% at four residues against 0.12% at five, the opposite direction to every other column in the paper. A tetrapeptide packs the same substituents around a twelve-membered ring, so its atoms start closer together and the refinement has less room to separate them. And a low rate buys nothing on its own. AvgFlow$_{\text{DiT}}$ is at or near zero at every size and still trails RINGER on ring W_1, which clashes in 17% to 27% of conformers. Appendix L takes that apart.

Table A7: Coverage and matching by residue count, recall only. Bold is best per column and panel, underline second best.

	COV-R ↑				MAT-R ↓			
Method	**4-mer**	**5-mer**	**6-mer**	**All**	**4-mer**	**5-mer**	**6-mer**	**All**
Ring; coverage threshold 0.10 Å								
ETKDGv3	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.500_{\pm 0.001}$	$0.553_{\pm 0.002}$	$0.839_{\pm 0.007}$	$0.559_{\pm 0.001}$
+ MMFF94	$56.0_{\pm 0.3}$	$1.2_{\pm 0.2}$	$0.0_{\pm 0.0}$	$30.3_{\pm 0.2}$	$0.177_{\pm 0.001}$	$0.330_{\pm 0.002}$	$0.565_{\pm 0.003}$	$0.277_{\pm 0.002}$
Torsional Diffusion	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.513_{\pm 0.005}$	$0.550_{\pm 0.001}$	$0.807_{\pm 0.006}$	$0.561_{\pm 0.003}$
RINGER	$\mathbf{88.5}_{\pm 0.7}$	$\underline{79.3}_{\pm 1.1}$	$\mathbf{58.3}_{\pm 2.5}$	$\underline{81.7}_{\pm 0.4}$	$\mathbf{0.070}_{\pm 0.003}$	$\underline{0.108}_{\pm 0.006}$	$\underline{0.241}_{\pm 0.027}$	$\mathbf{0.104}_{\pm 0.001}$
MCF-S	$69.3_{\pm 4.4}$	$15.5_{\pm 7.5}$	$0.4_{\pm 0.5}$	$42.4_{\pm 5.0}$	$0.119_{\pm 0.004}$	$0.205_{\pm 0.011}$	$0.320_{\pm 0.015}$	$0.173_{\pm 0.007}$
ET-Flow-*SO(3)*	$81.4_{\pm 1.0}$	$42.6_{\pm 3.1}$	$7.3_{\pm 1.5}$	$59.0_{\pm 1.7}$	$0.094_{\pm 0.001}$	$0.175_{\pm 0.001}$	$0.322_{\pm 0.014}$	$0.149_{\pm 0.003}$
AGDiff	$51.7_{\pm 1.6}$	$4.1_{\pm 1.0}$	$0.1_{\pm 0.1}$	$29.0_{\pm 0.6}$	$0.179_{\pm 0.003}$	$0.369_{\pm 0.003}$	$0.570_{\pm 0.009}$	$0.292_{\pm 0.000}$
AvgFlow$_{\text{DiT}}$	$78.7_{\pm 1.2}$	$58.4_{\pm 0.7}$	$22.2_{\pm 3.2}$	$64.9_{\pm 0.8}$	$0.144_{\pm 0.008}$	$0.207_{\pm 0.009}$	$0.305_{\pm 0.011}$	$0.185_{\pm 0.008}$
CycloFlow	$\underline{88.0}_{\pm 0.4}$	$\mathbf{81.8}_{\pm 0.2}$	$\underline{57.8}_{\pm 2.0}$	$\mathbf{82.3}_{\pm 0.5}$	$\underline{0.083}_{\pm 0.002}$	$\mathbf{0.101}_{\pm 0.001}$	$\mathbf{0.236}_{\pm 0.021}$	$\underline{0.108}_{\pm 0.004}$
All-atom; coverage threshold 0.75 Å								
ETKDGv3	$2.3_{\pm 0.1}$	$0.2_{\pm 0.1}$	$0.1_{\pm 0.1}$	$1.3_{\pm 0.1}$	$1.314_{\pm 0.002}$	$1.496_{\pm 0.002}$	$2.047_{\pm 0.007}$	$1.465_{\pm 0.001}$
+ MMFF94	$31.7_{\pm 0.2}$	$9.0_{\pm 0.6}$	$0.8_{\pm 0.3}$	$20.1_{\pm 0.2}$	$0.990_{\pm 0.000}$	$1.255_{\pm 0.010}$	$1.725_{\pm 0.007}$	$1.170_{\pm 0.003}$
Torsional Diffusion	$1.3_{\pm 0.4}$	$0.2_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.8_{\pm 0.2}$	$1.337_{\pm 0.019}$	$1.498_{\pm 0.022}$	$1.937_{\pm 0.014}$	$1.465_{\pm 0.018}$
RINGER	$\underline{82.3}_{\pm 0.2}$	$\underline{73.5}_{\pm 0.7}$	$\underline{60.3}_{\pm 2.2}$	$\underline{76.6}_{\pm 0.1}$	$\underline{0.456}_{\pm 0.006}$	$\underline{0.588}_{\pm 0.009}$	$\underline{0.901}_{\pm 0.049}$	$\underline{0.555}_{\pm 0.004}$
MCF-S	$66.3_{\pm 3.1}$	$39.9_{\pm 5.6}$	$13.7_{\pm 3.9}$	$50.8_{\pm 4.1}$	$0.653_{\pm 0.037}$	$0.943_{\pm 0.047}$	$1.254_{\pm 0.051}$	$0.826_{\pm 0.042}$
ET-Flow-*SO(3)*	$67.1_{\pm 0.5}$	$40.4_{\pm 1.5}$	$11.7_{\pm 0.6}$	$51.2_{\pm 0.8}$	$0.625_{\pm 0.007}$	$0.929_{\pm 0.010}$	$1.302_{\pm 0.023}$	$0.812_{\pm 0.008}$
AGDiff	$11.4_{\pm 0.5}$	$1.9_{\pm 0.5}$	$0.0_{\pm 0.0}$	$6.7_{\pm 0.2}$	$1.179_{\pm 0.010}$	$1.516_{\pm 0.012}$	$1.902_{\pm 0.025}$	$1.383_{\pm 0.005}$
AvgFlow$_{\text{DiT}}$	$69.3_{\pm 1.5}$	$57.8_{\pm 0.1}$	$33.7_{\pm 2.8}$	$61.1_{\pm 0.8}$	$0.653_{\pm 0.022}$	$0.885_{\pm 0.018}$	$1.200_{\pm 0.014}$	$0.799_{\pm 0.016}$
CycloFlow	$\mathbf{84.9}_{\pm 0.4}$	$\mathbf{79.9}_{\pm 1.2}$	$\mathbf{63.7}_{\pm 3.1}$	$\mathbf{80.6}_{\pm 0.0}$	$\mathbf{0.376}_{\pm 0.009}$	$\mathbf{0.465}_{\pm 0.009}$	$\mathbf{0.831}_{\pm 0.039}$	$\mathbf{0.461}_{\pm 0.004}$

Table A8: Coverage and matching by residue count, precision only. Bold is best per column and panel, underline second best.

	COV-P ↑				MAT-P ↓			
Method	**4-mer**	**5-mer**	**6-mer**	**All**	**4-mer**	**5-mer**	**6-mer**	**All**
Ring; coverage threshold 0.10 Å								
ETKDGv3	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.656_{\pm 0.001}$	$0.779_{\pm 0.000}$	$1.189_{\pm 0.000}$	$0.762_{\pm 0.000}$
+ MMFF94	$10.2_{\pm 0.0}$	$0.2_{\pm 0.0}$	$0.0_{\pm 0.0}$	$5.5_{\pm 0.0}$	$0.586_{\pm 0.001}$	$0.729_{\pm 0.000}$	$1.166_{\pm 0.001}$	$0.705_{\pm 0.001}$
Torsional Diffusion	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.645_{\pm 0.000}$	$0.766_{\pm 0.001}$	$1.174_{\pm 0.001}$	$0.751_{\pm 0.000}$
RINGER	$\underline{67.6}_{\pm 0.7}$	$\underline{51.8}_{\pm 3.1}$	$\underline{30.8}_{\pm 3.2}$	$\underline{57.7}_{\pm 1.3}$	$\underline{0.340}_{\pm 0.016}$	$\underline{0.589}_{\pm 0.003}$	$\underline{0.801}_{\pm 0.007}$	$\underline{0.482}_{\pm 0.007}$
MCF-S	$36.8_{\pm 3.5}$	$2.1_{\pm 1.4}$	$0.0_{\pm 0.0}$	$20.4_{\pm 2.3}$	$0.426_{\pm 0.013}$	$0.613_{\pm 0.018}$	$1.025_{\pm 0.017}$	$0.563_{\pm 0.015}$
ET-Flow-*SO(3)*	$42.8_{\pm 2.7}$	$7.9_{\pm 1.0}$	$0.5_{\pm 0.2}$	$25.6_{\pm 1.8}$	$0.398_{\pm 0.008}$	$0.614_{\pm 0.004}$	$1.048_{\pm 0.014}$	$0.551_{\pm 0.002}$
AGDiff	$4.8_{\pm 0.3}$	$0.3_{\pm 0.0}$	$0.0_{\pm 0.0}$	$2.7_{\pm 0.2}$	$0.713_{\pm 0.007}$	$0.873_{\pm 0.010}$	$1.254_{\pm 0.012}$	$0.834_{\pm 0.008}$
AvgFlow$_{\text{DiT}}$	$53.2_{\pm 1.4}$	$18.8_{\pm 1.2}$	$3.3_{\pm 0.6}$	$35.3_{\pm 1.2}$	$0.370_{\pm 0.016}$	$0.516_{\pm 0.013}$	$0.951_{\pm 0.022}$	$0.490_{\pm 0.014}$
CycloFlow	$\mathbf{69.5}_{\pm 2.8}$	$\mathbf{65.8}_{\pm 0.6}$	$\mathbf{39.4}_{\pm 1.7}$	$\mathbf{64.6}_{\pm 1.8}$	$\mathbf{0.186}_{\pm 0.010}$	$\mathbf{0.189}_{\pm 0.004}$	$\mathbf{0.461}_{\pm 0.016}$	$\mathbf{0.220}_{\pm 0.005}$
All-atom; coverage threshold 0.75 Å								
ETKDGv3	$0.7_{\pm 0.0}$	$0.1_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.4_{\pm 0.0}$	$1.565_{\pm 0.001}$	$1.934_{\pm 0.000}$	$2.701_{\pm 0.001}$	$1.830_{\pm 0.000}$
+ MMFF94	$7.6_{\pm 0.1}$	$1.8_{\pm 0.0}$	$0.1_{\pm 0.0}$	$4.7_{\pm 0.1}$	$1.488_{\pm 0.002}$	$1.881_{\pm 0.001}$	$2.689_{\pm 0.003}$	$1.768_{\pm 0.002}$
Torsional Diffusion	$0.4_{\pm 0.1}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.2_{\pm 0.1}$	$1.635_{\pm 0.017}$	$1.968_{\pm 0.007}$	$2.715_{\pm 0.010}$	$1.880_{\pm 0.010}$
RINGER	$\underline{64.9}_{\pm 0.7}$	$\underline{50.7}_{\pm 1.2}$	$\underline{33.0}_{\pm 1.4}$	$\underline{56.2}_{\pm 0.2}$	$\underline{0.835}_{\pm 0.021}$	$\underline{1.207}_{\pm 0.012}$	$\underline{1.674}_{\pm 0.006}$	$\underline{1.065}_{\pm 0.007}$
MCF-S	$34.6_{\pm 2.8}$	$11.3_{\pm 2.8}$	$2.6_{\pm 0.7}$	$22.7_{\pm 2.6}$	$1.214_{\pm 0.044}$	$1.711_{\pm 0.042}$	$2.442_{\pm 0.032}$	$1.534_{\pm 0.042}$
ET-Flow-*SO(3)*	$35.3_{\pm 2.0}$	$11.7_{\pm 0.5}$	$1.7_{\pm 0.0}$	$23.1_{\pm 1.3}$	$1.170_{\pm 0.027}$	$1.717_{\pm 0.004}$	$2.500_{\pm 0.017}$	$1.520_{\pm 0.012}$
AGDiff	$1.6_{\pm 0.1}$	$0.3_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.9_{\pm 0.0}$	$1.922_{\pm 0.014}$	$2.322_{\pm 0.017}$	$2.967_{\pm 0.012}$	$2.186_{\pm 0.013}$
AvgFlow$_{\text{DiT}}$	$44.3_{\pm 1.7}$	$25.3_{\pm 1.4}$	$7.7_{\pm 1.0}$	$33.3_{\pm 1.4}$	$1.103_{\pm 0.041}$	$1.472_{\pm 0.024}$	$2.264_{\pm 0.039}$	$1.370_{\pm 0.028}$
CycloFlow	$\mathbf{77.0}_{\pm 0.9}$	$\mathbf{69.4}_{\pm 0.6}$	$\mathbf{43.9}_{\pm 1.5}$	$\mathbf{70.4}_{\pm 0.5}$	$\mathbf{0.527}_{\pm 0.019}$	$\mathbf{0.628}_{\pm 0.008}$	$\mathbf{1.246}_{\pm 0.036}$	$\mathbf{0.648}_{\pm 0.011}$

Table A9: Ablations by residue count, recall only. Every row changes one stage of the full pipeline, named in the first column. Bold is best per column and panel, underline second best.

	COV-R ↑				MAT-R ↓			
Configuration	**4-mer**	**5-mer**	**6-mer**	**All**	**4-mer**	**5-mer**	**6-mer**	**All**
Ring; coverage threshold 0.10 Å								
CycloFlow	88.0$_{\pm 0.4}$	81.8$_{\pm 0.2}$	57.8$_{\pm 2.0}$	82.3$_{\pm 0.5}$	0.083$_{\pm 0.002}$	0.101$_{\pm 0.001}$	0.236$_{\pm 0.021}$	0.108$_{\pm 0.004}$
First stage only	**89.3**$_{\pm 0.3}$	81.8$_{\pm 0.2}$	55.6$_{\pm 1.9}$	**82.7**$_{\pm 0.3}$	**0.077**$_{\pm 0.003}$	**0.092**$_{\pm 0.001}$	**0.225**$_{\pm 0.004}$	**0.100**$_{\pm 0.001}$
No steric refinement	88.0$_{\pm 0.5}$	**81.9**$_{\pm 0.2}$	58.1$_{\pm 2.0}$	82.3$_{\pm 0.5}$	0.083$_{\pm 0.003}$	0.101$_{\pm 0.001}$	0.236$_{\pm 0.021}$	0.108$_{\pm 0.004}$
No closure solve	71.2$_{\pm 1.2}$	73.8$_{\pm 0.3}$	48.9$_{\pm 1.8}$	69.4$_{\pm 0.5}$	0.128$_{\pm 0.001}$	0.120$_{\pm 0.001}$	0.251$_{\pm 0.020}$	0.140$_{\pm 0.003}$
SLSQP closure	83.8$_{\pm 0.6}$	78.1$_{\pm 0.5}$	**58.7**$_{\pm 2.6}$	78.8$_{\pm 0.3}$	0.103$_{\pm 0.003}$	0.108$_{\pm 0.003}$	0.253$_{\pm 0.024}$	0.123$_{\pm 0.002}$
All-atom; coverage threshold 0.75 Å								
CycloFlow	84.9$_{\pm 0.4}$	79.9$_{\pm 1.2}$	63.7$_{\pm 3.1}$	80.6$_{\pm 0.0}$	0.376$_{\pm 0.009}$	0.465$_{\pm 0.009}$	0.831$_{\pm 0.039}$	0.461$_{\pm 0.004}$
First stage only	**86.8**$_{\pm 0.5}$	**82.2**$_{\pm 0.4}$	**64.8**$_{\pm 2.2}$	**82.5**$_{\pm 0.3}$	**0.361**$_{\pm 0.004}$	**0.447**$_{\pm 0.005}$	**0.799**$_{\pm 0.019}$	**0.443**$_{\pm 0.003}$
No steric refinement	84.9$_{\pm 0.4}$	79.9$_{\pm 1.2}$	63.6$_{\pm 3.1}$	80.6$_{\pm 0.1}$	0.376$_{\pm 0.009}$	0.465$_{\pm 0.009}$	0.832$_{\pm 0.039}$	0.462$_{\pm 0.004}$
No closure solve	83.8$_{\pm 0.3}$	79.6$_{\pm 1.1}$	63.7$_{\pm 3.1}$	79.9$_{\pm 0.1}$	0.456$_{\pm 0.008}$	0.481$_{\pm 0.008}$	0.839$_{\pm 0.043}$	0.511$_{\pm 0.005}$
SLSQP closure	83.5$_{\pm 0.5}$	77.6$_{\pm 0.7}$	61.2$_{\pm 2.6}$	78.8$_{\pm 0.4}$	0.428$_{\pm 0.007}$	0.505$_{\pm 0.007}$	0.863$_{\pm 0.052}$	0.507$_{\pm 0.005}$

Table A10: Ablations by residue count, precision only. Every row changes one stage of the full pipeline, named in the first column. Bold is best per column and panel, underline second best.

	COV-P ↑				MAT-P ↓			
Configuration	**4-mer**	**5-mer**	**6-mer**	**All**	**4-mer**	**5-mer**	**6-mer**	**All**
Ring; coverage threshold 0.10 Å								
CycloFlow	69.5$_{\pm 2.8}$	**65.8**$_{\pm 0.6}$	39.4$_{\pm 1.7}$	**64.6**$_{\pm 1.8}$	**0.186**$_{\pm 0.010}$	**0.189**$_{\pm 0.004}$	**0.461**$_{\pm 0.016}$	**0.220**$_{\pm 0.005}$
First stage only	64.7$_{\pm 2.9}$	57.3$_{\pm 0.7}$	30.7$_{\pm 0.8}$	58.0$_{\pm 1.8}$	0.212$_{\pm 0.007}$	0.220$_{\pm 0.004}$	0.501$_{\pm 0.015}$	0.249$_{\pm 0.003}$
No steric refinement	**69.6**$_{\pm 2.8}$	**65.8**$_{\pm 0.7}$	**39.5**$_{\pm 1.7}$	**64.6**$_{\pm 1.8}$	0.191$_{\pm 0.011}$	**0.189**$_{\pm 0.004}$	**0.461**$_{\pm 0.016}$	0.223$_{\pm 0.006}$
No closure solve	7.9$_{\pm 0.1}$	36.0$_{\pm 0.7}$	29.3$_{\pm 1.7}$	20.2$_{\pm 0.5}$	0.333$_{\pm 0.012}$	0.239$_{\pm 0.003}$	0.484$_{\pm 0.017}$	0.319$_{\pm 0.007}$
SLSQP closure	30.6$_{\pm 0.7}$	38.6$_{\pm 0.2}$	35.6$_{\pm 1.3}$	34.0$_{\pm 0.5}$	0.436$_{\pm 0.010}$	0.585$_{\pm 0.003}$	0.798$_{\pm 0.013}$	0.531$_{\pm 0.007}$
All-atom; coverage threshold 0.75 Å								
CycloFlow	**77.0**$_{\pm 0.9}$	**69.4**$_{\pm 0.6}$	**43.9**$_{\pm 1.5}$	**70.4**$_{\pm 0.5}$	**0.527**$_{\pm 0.019}$	**0.628**$_{\pm 0.008}$	**1.246**$_{\pm 0.036}$	**0.648**$_{\pm 0.011}$
First stage only	71.8$_{\pm 1.1}$	63.0$_{\pm 0.6}$	37.8$_{\pm 0.8}$	64.7$_{\pm 0.8}$	0.601$_{\pm 0.018}$	0.724$_{\pm 0.011}$	1.357$_{\pm 0.026}$	0.734$_{\pm 0.011}$
No steric refinement	**77.0**$_{\pm 0.9}$	**69.4**$_{\pm 0.6}$	**43.9**$_{\pm 1.5}$	**70.4**$_{\pm 0.6}$	0.530$_{\pm 0.020}$	**0.628**$_{\pm 0.008}$	1.247$_{\pm 0.036}$	0.650$_{\pm 0.011}$
No closure solve	73.7$_{\pm 1.0}$	69.1$_{\pm 0.5}$	43.7$_{\pm 1.5}$	68.5$_{\pm 0.6}$	0.651$_{\pm 0.018}$	0.652$_{\pm 0.007}$	1.255$_{\pm 0.037}$	0.723$_{\pm 0.011}$
SLSQP closure	62.4$_{\pm 0.9}$	52.5$_{\pm 0.4}$	36.1$_{\pm 1.5}$	55.8$_{\pm 0.7}$	0.952$_{\pm 0.019}$	1.209$_{\pm 0.009}$	1.672$_{\pm 0.033}$	1.128$_{\pm 0.015}$

Table A11: Per-size W_1 and clash rate (Appendix G). Coverage and matching are in Appendix J. Bold is best, underline second best.

	Ring W_1 (Å) ↓			All-atom W_1 (Å) ↓			Clash (%) ↓		
Method	4-mer	5-mer	6-mer	4-mer	5-mer	6-mer	4-mer	5-mer	6-mer
ETKDGv3	0.773$_{\pm 0.004}$	0.893$_{\pm 0.002}$	1.320$_{\pm 0.002}$	1.823$_{\pm 0.003}$	2.272$_{\pm 0.006}$	3.030$_{\pm 0.014}$	1.87$_{\pm 0.07}$	0.57$_{\pm 0.01}$	0.25$_{\pm 0.03}$
+ MMFF94	0.696$_{\pm 0.001}$	0.846$_{\pm 0.004}$	1.274$_{\pm 0.001}$	1.805$_{\pm 0.006}$	2.237$_{\pm 0.011}$	3.025$_{\pm 0.015}$	**0.00**$_{\pm 0.00}$	**0.00**$_{\pm 0.00}$	**0.00**$_{\pm 0.00}$
Torsional Diffusion	0.760$_{\pm 0.004}$	0.881$_{\pm 0.004}$	1.291$_{\pm 0.006}$	1.865$_{\pm 0.012}$	2.283$_{\pm 0.016}$	3.024$_{\pm 0.010}$	22.14$_{\pm 6.57}$	27.79$_{\pm 7.50}$	39.16$_{\pm 8.65}$
RINGER	0.299$_{\pm 0.008}$	0.603$_{\pm 0.027}$	0.868$_{\pm 0.018}$	1.025$_{\pm 0.009}$	1.424$_{\pm 0.055}$	1.961$_{\pm 0.061}$	17.04$_{\pm 0.60}$	25.80$_{\pm 2.42}$	27.08$_{\pm 3.51}$
MCF-S	0.515$_{\pm 0.015}$	0.753$_{\pm 0.020}$	1.148$_{\pm 0.007}$	1.491$_{\pm 0.063}$	2.077$_{\pm 0.044}$	2.791$_{\pm 0.006}$	22.23$_{\pm 16.19}$	28.47$_{\pm 20.42}$	43.07$_{\pm 22.77}$
ET-Flow-*SO(3)*	0.479$_{\pm 0.022}$	0.765$_{\pm 0.005}$	1.150$_{\pm 0.027}$	1.485$_{\pm 0.023}$	2.094$_{\pm 0.015}$	2.780$_{\pm 0.028}$	1.12$_{\pm 0.24}$	6.88$_{\pm 0.73}$	20.73$_{\pm 2.32}$
AGDiff	0.806$_{\pm 0.005}$	0.990$_{\pm 0.007}$	1.353$_{\pm 0.012}$	2.164$_{\pm 0.023}$	2.623$_{\pm 0.029}$	3.218$_{\pm 0.031}$	0.69$_{\pm 0.05}$	6.21$_{\pm 0.37}$	21.71$_{\pm 1.16}$
AvgFlow$_{\text{DiT}}$	0.448$_{\pm 0.043}$	0.644$_{\pm 0.018}$	1.104$_{\pm 0.059}$	1.343$_{\pm 0.046}$	1.789$_{\pm 0.049}$	2.625$_{\pm 0.060}$	**0.00**$_{\pm 0.00}$	0.02$_{\pm 0.01}$	0.06$_{\pm 0.04}$
CycloFlow	**0.158**$_{\pm 0.011}$	**0.228**$_{\pm 0.015}$	**0.522**$_{\pm 0.039}$	**0.757**$_{\pm 0.014}$	**0.872**$_{\pm 0.014}$	**1.547**$_{\pm 0.067}$	2.37$_{\pm 0.74}$	0.12$_{\pm 0.03}$	0.39$_{\pm 0.17}$

Table A12: Clashes and xTB ensemble properties. Property columns are the median error in the Boltzmann-averaged quantity after the xTB relaxation of Appendix H. The three clash columns overlap, since a conformer with contacts of two kinds is counted in both, so they do not sum to the total in Table 1. Bold is best per column, underline second best.

	xTB ensemble property error ↓				Clash rate by contact type (%) ↓		
Method	E **(kcal/mol)**	μ **(D)**	$\Delta\varepsilon$ **(kcal/mol)**	$E_{\min}$ **(kcal/mol)**	**Ring–ring**	**Ring–side**	**Side–side**
RINGER	$0.401_{\pm 0.027}$	$0.573_{\pm 0.021}$	$0.640_{\pm 0.022}$	$0.354_{\pm 0.048}$	$15.9_{\pm 0.4}$	$18.9_{\pm 0.9}$	$20.1_{\pm 1.1}$
ET-Flow-*SO(3)*	$1.346_{\pm 0.057}$	$1.224_{\pm 0.074}$	$1.447_{\pm 0.050}$	$1.385_{\pm 0.061}$	$\mathbf{0.0}_{\pm 0.0}$	$\underline{0.6}_{\pm 0.1}$	$5.3_{\pm 0.5}$
AvgFlow$_{\mathrm{DiT}}$	$\underline{0.333}_{\pm 0.019}$	$\underline{0.544}_{\pm 0.045}$	$\underline{0.588}_{\pm 0.033}$	$\underline{0.180}_{\pm 0.030}$	$\mathbf{0.0}_{\pm 0.0}$	$\mathbf{0.0}_{\pm 0.0}$	$\mathbf{0.0}_{\pm 0.0}$
CYCLOFLOW	$\mathbf{0.266}_{\pm 0.013}$	$\mathbf{0.494}_{\pm 0.014}$	$\mathbf{0.522}_{\pm 0.027}$	$\mathbf{0.038}_{\pm 0.003}$	$0.9_{\pm 0.3}$	$1.3_{\pm 0.4}$	$\underline{0.5}_{\pm 0.2}$
No steric refinement	$0.267_{\pm 0.013}$	$0.504_{\pm 0.018}$	$0.518_{\pm 0.034}$	$0.037_{\pm 0.005}$	$1.4_{\pm 0.5}$	$2.7_{\pm 0.7}$	$\underline{4.5}_{\pm 1.0}$

L ADDITIONAL RESULTS: XTB ENSEMBLE PROPERTIES AND CLASHES

Ensemble matching rewards a conformer for being close to a reference and does not ask whether it behaves like one. We relax every conformer with GFN2-xTB under one protocol for all methods (Appendix H) and compare four Boltzmann-averaged quantities of the relaxed ensemble against the reference, alongside the clash rate.

CYCLOFLOW is closest on all four properties, most of all in $E_{\min}$, where the error falls from AvgFlow$_{\mathrm{DiT}}$'s 0.180 to 0.038 kcal/mol. Clashes tell a different story. AvgFlow$_{\mathrm{DiT}}$ and ET-Flow-*SO(3)* are the cleanest there and sit five and thirty-six times our $E_{\min}$ error, so a low clash rate certifies that no two atoms are too close, not that the ensemble sits in the right region of the energy landscape. Our own rate is 1.4% against RINGER's 21.3%, and the improvement is in the side chains, where contacts fall from 20.1% to 0.5%. Since the steric refinement carries a repulsion term we give both sides of it, and before it CYCLOFLOW clashes in $4.6 \pm 0.8\%$ of conformers, already under a quarter of RINGER's rate.

Where RINGER's clashes come from we can answer on our own model. They are not a bond length effect, correlating with a conformer's ring bond length error at only 0.12. They track the ring instead, the median clashing RINGER conformer sitting 1.79 Å from its nearest reference where a clean one sits at 0.02, a rank correlation of 0.64. Swapping our closure solve for RINGER's on our own samples reproduces the rate (Section 4.2), which places those overlaps in the structure rather than in local contacts.

M ADDITIONAL ANALYSIS: COUPLING BETWEEN RING TORSIONS

The two measures of Section 5.1 read different parts of the same table, so one can be good while the other is bad. The *coupling strength* divides the generated mutual information by the reference's, so it reads how tightly the pair moves together and not which wells it moves between. The *combination error* is the Jensen–Shannon divergence between the two tables of combination shares, so those shares are most of what it charges for. A model that puts its mass in the wrong wells and pairs them tightly scores well on the first and badly on the second, and a model that finds the right wells and pairs them loosely scores the other way round (Figure A4).

This is the pattern the 4-mer column of Section 5.1 holds, where CYCLOFLOW records the lowest combination error of the four methods while holding the least coupling.

N ADDITIONAL ANALYSIS: WHICH ROTAMER A SIDE CHAIN OCCUPIES

Section 4.3 pools χ_1 over the test set, which averages every molecule's occupancy together, so over-using a rotamer in one molecule and under-using it in another scores as neither. The joint measurement of Section 5.1 does not carry over, since side chains give a reference coupling of 0.021 against 0.110, 0.099 and 0.094 on the ring.

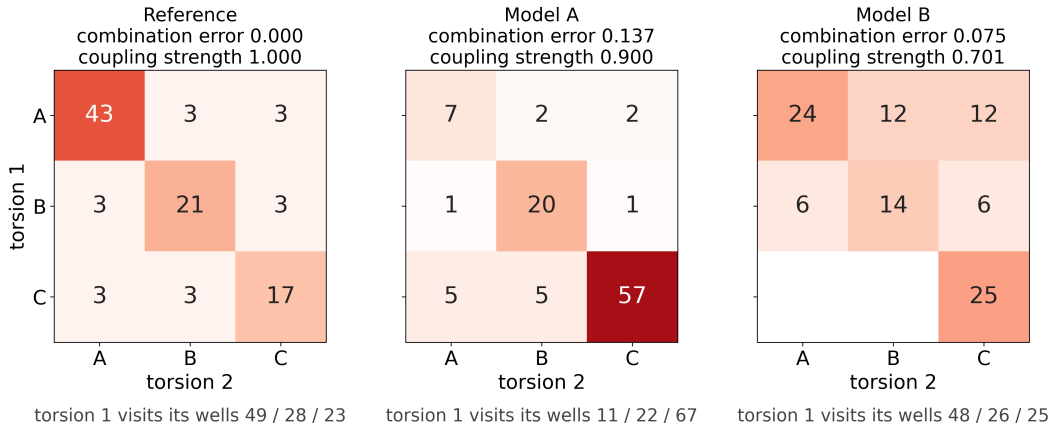

Figure A4: The two measures can disagree. Three constructed tables over the wells of one torsion pair. Model A keeps sharp, well separated rows, so knowing torsion 1 fixes torsion 2 and its coupling strength is the higher of the two. It also sits in the wrong rows, visiting torsion 1's wells 11, 22 and 67 percent of the time against the reference's 49, 28 and 23, so its combination error is the worse of the two. Model B carries the reference's own row shares and blurrier rows, and scores the other way round. Nothing here is measured.

Table A13: Where a generated side chain sits, read against the wells that molecule's own reference occupies. **Outside** is the share of generated conformers whose χ_1 is further than $60°$ from every one of those wells, a rotamer the reference never uses. **Weights** is the total variation distance between the generated and reference weights over the wells the reference does use. Both are read over 20 conformers a side, and the floor scores one half of a reference ensemble against wells taken from the other half. Bold is best per column, underline second best.

	4-mer (529)		**5-mer** (347)		**6-mer** (120)	
Method	**Outside** (%) ↓ floor 1.06	**Weights** ↓ floor 0.126	**Outside** (%) ↓ floor 1.36	**Weights** ↓ floor 0.115	**Outside** (%) ↓ floor 1.26	**Weights** ↓ floor 0.107
RINGER	$\underline{2.09}_{\pm 0.07}$	$\underline{0.185}_{\pm 0.004}$	$\underline{3.34}_{\pm 0.11}$	$\underline{0.209}_{\pm 0.002}$	$\mathbf{4.50}_{\pm 0.10}$	$\underline{0.245}_{\pm 0.006}$
ET-Flow-*SO(3)*	$7.31_{\pm 0.23}$	$0.224_{\pm 0.001}$	$9.86_{\pm 0.49}$	$0.257_{\pm 0.001}$	$12.34_{\pm 0.17}$	$0.268_{\pm 0.004}$
AvgFlow$_{\text{DiT}}$	$6.65_{\pm 0.36}$	$0.235_{\pm 0.005}$	$9.38_{\pm 0.28}$	$0.260_{\pm 0.001}$	$11.56_{\pm 0.33}$	$0.273_{\pm 0.003}$
CycloFlow	$\mathbf{1.63}_{\pm 0.05}$	$\mathbf{0.177}_{\pm 0.001}$	$\mathbf{2.82}_{\pm 0.12}$	$\mathbf{0.204}_{\pm 0.003}$	$\underline{4.79}_{\pm 0.23}$	$\mathbf{0.241}_{\pm 0.008}$

We therefore read one side chain at a time, against the wells its own reference ensemble occupies, which are the staggered rotamers (Figure A5a). A generated ensemble can leave those wells for a rotamer the reference never uses, or stay inside them and weight them wrongly, and Table A13 separates the two over 20 conformers a side. Both have a floor, since wells taken from one half of a reference ensemble do not cover the other half exactly either.

Leaving the wells is rare for the methods that generate internal coordinates. CycloFlow is at 1.63% on 4-mers against a floor of 1.06, where the Cartesian methods are at 7.31% and 6.65% and reach 12.34% and 11.56% on 6-mers. Weighting them is where no method reaches the floor. CycloFlow is closest at every residue count and still sits at 0.177 against 0.126 on 4-mers and 0.241 against 0.107 on 6-mers. Figure A5b shows both on one side chain. Every method peaks where the reference peaks, and AvgFlow$_{\text{DiT}}$ gives 41% to a well the reference gives 18%.

Two limits belong with this. The measurement reads the first side-chain torsion angle only, and the wells come from forty reference conformers, so a rotamer the reference never sampled counts against a model that finds it.

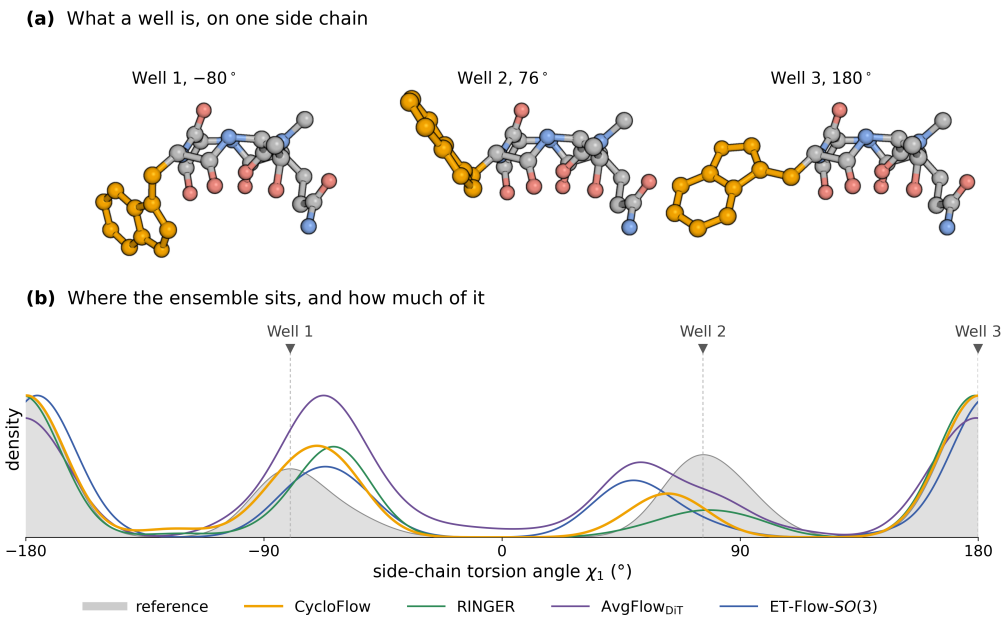

Figure A5: Rotamers of one side chain. (a) Three reference conformers of one test molecule, one at each well of the side chain drawn in color. They are superimposed on the ring and chosen so that the rest of the molecule matches, so what differs between the panels is that side chain. (b) The distribution those wells come from with the share of each ensemble that sits in each well.

O Additional analysis: What counts as a closed ring

Section 5.2 separates the methods at 0.0057 Å. That tolerance comes from the references' own reproducibility rather than from chemistry. Every CREMP conformer is optimized at the same level, so a bond sits at its own minimum and barely moves from one conformer of a molecule to the next, which leaves the closing bond a within-molecule spread of 0.00147 Å. A tolerance four times that measures agreement with an optimized structure rather than whether the ring closed.

The other scale is energy. We hold the closing bond, or the bond angle at the same seam, away from its relaxed value and minimize MMFF94 over every other coordinate, on 120 test molecules. The strain that survives that minimization converts a deviation into k_BT, the unit panel (b) already uses. One k_BT costs 0.0955 Å on the bond and $4.55°$ on the angle. Ten k_BT, the limit of panel (b), costs 0.300 Å and $13.9°$. The scan relaxes the rest of the molecule, which is the right question for whether a molecule would adopt the conformer. Holding the other coordinates fixed would place both thresholds lower.

Read against that scale, the seam is not where any method fails (Table A14). Every one of them keeps at least 99.7% of its conformers inside $1\,k_BT$ on the bond, and the worst median, ET-Flow-*SO(3)* at 0.0118 Å, is $0.01\,k_BT$. What Figure 4a separates at 0.0057 Å is how closely each method reproduces an optimized reference, not whether its rings are closed.

The same table separates our two closure components, which act on different coordinates. The solve turns six ring coordinates so that the chain returns to where it started, and the steric refinement then moves atoms off one another under a restraint holding the closing bond at the training-split mean. Removing the refinement leaves the bond median at 0.0137 Å, which is $0.02\,k_BT$, and 96.7% of conformers inside $1\,k_BT$, so the ring still closes. Removing the solve leaves the bond alone, since the refinement still holds it, and breaks the angle instead. Its median reaches $11.84°$, which is $6.4\,k_BT$ of strain, and only 21.4% of conformers stay inside $1\,k_BT$. That is a ring that did not close, and it is what the ring COV-F1 of 31.3% in Table 1 records.

Table A14: The two seam coordinates, pooled over three seeds. Medians are the deviation from that molecule's own reference. The share columns count conformers inside the strain threshold, 0.0955 Å on the bond and $4.55°$ on the angle. The last row is the references' own within-molecule spread, which is the floor of the two median columns.

	Median deviation		Within $1\,k_BT$ (%)	
Method	**Bond (Å)**	**Angle (°)**	**Bond**	**Angle**
RINGER	0.0021	1.08	100.0	75.9
ET-Flow-*SO(3)*	0.0118	1.99	99.7	80.9
AvgFlow$_{\mathrm{DiT}}$	0.0071	1.16	100.0	87.9
CycloFlow	0.0021	1.06	100.0	89.2
No steric refinement	0.0137	1.45	96.7	85.5
No closure solve	0.0024	11.84	99.4	21.4
SLSQP closure	0.0021	2.00	100.0	72.2
Reference spread	0.00147	0.823	–	–

P Limitations

Ring size is a real difficulty, measured against a truncated view of it. Every method degrades with residue count, our own ring coverage falling from 77.7% at four residues to 72.9% at five and 46.8% at six. Whether that is the problem growing or the model weakening is a question this benchmark cannot settle, and one statistic inside it points the wrong way. Among the forty kept conformers the effective number, which discounts those carrying little Boltzmann weight, is 14.8, 13.9 and 14.0 across the three sizes, a flatness that would put the whole fall on the model. It is an artifact of the cap. The CREST ensembles those forty come from hold a median of 511, 856 and 1,006 unique conformers, so the cap spans 7.8% of a tetrapeptide's ensemble and 4.0% of a hexapeptide's, and the distribution a generator has to match roughly doubles over this range.

Two explanations can be ruled out. The fall is not a shortage of degrees of freedom, since within a residue count the molecules with more rotatable ring bonds do better, not worse, and 4-mers with six reach 80.6% where those with four reach 31.4%. Nor is it error accumulating along the reconstruction, since the per-atom deviation is flat in walk position and flattest of all at six residues. What tracks it is the coupling. Within a residue count the per-molecule combination error and ring coverage correlate at -0.64, -0.67 and -0.64, and the quarter of 6-mers whose coupling we reproduce best still reaches only 69.3% against 90.4% for the same quarter of 4-mers. That is an association rather than a cause, since a shortage of 6-mer training molecules would produce both.

Side chains reach the right rotamers in the wrong proportions. The error is in the proportions rather than the placement. Leaving the reference's own rotamers is rare for us, where weighting them reaches no method's floor (Appendix N). Within a residue count, the number of side-chain heavy atoms correlates at 0.182, 0.208 and 0.116 with how hard the target is on its own terms, measured by scoring each CREMP ensemble against a held-out half of itself, and at 0.276, 0.253 and 0.330 with our own all-atom matching error, so the gap is ours to close and it is widest on the largest rings.

The reference is a level of theory, not an experiment. CREMP's conformers come from CREST followed by GFN2-xTB with an implicit chloroform model (Grambow et al., 2024), and every number in this paper is agreement with that. A semiempirical Hamiltonian and a continuum solvent are approximations, so a generator that matched these ensembles exactly would inherit whatever they get wrong, and the energies of Appendix L are recomputed at the same level rather than an independent one. What the benchmark establishes is that a generator can stand in for the search that produced the reference, at a cost of seconds rather than hours. Whether that reference is right for a given downstream question is a separate matter no amount of agreement here settles.

Only head-to-tail cycles, and the closing bond has to be named. CREMP holds head-to-tail cyclic peptides, so nothing here is evidence about side-chain-to-tail lactams, disulfides or any other way of closing a macrocycle. The method also takes the ring and one bond on it as given, since the reconstruction walks from three seed atoms, the closure residual is written where the walk returns, and the six solved coordinates are chosen relative to that bond. Nothing in that requires a backbone amide and any ring in the graph would serve, but the model has only ever been trained on one kind of closure. A peptide with two rings is outside the formulation altogether, since the walk and the residual are written for one.