CycPepFlow: Reconciling Ring Coupling and Stereochemical Fidelity in Cyclic Peptide Ensemble Generation

Anonymous authors
Paper under double-blind review

Abstract

Cyclic peptides are promising therapeutic scaffolds, but generating their conformational ensembles requires capturing long-range ring coupling while preserving local chemical geometry and stereochemistry. We introduce CycPepFlow, an all-atom Cartesian flow model that generates conformers directly from molecular graphs. CycPepFlow combines SO(3)-equivariant local interactions with gated global attention, while atomic chirality conditioning and signed-volume supervision promote tetrahedral stereochemical fidelity. An optional all-pair graph-geodesic attention bias (APG) incorporates pairwise topology into global communication. Evaluated on 1,000 held-out CREMP cyclic peptides containing four to six residues, CycPepFlow-L achieves the highest coverage F1 among the compared methods (76.13%), whereas CycPepFlow APG-L achieves the highest generated-set coverage precision (79.78%) and lowest generated-to-reference matching distance (0.419 Å). On the six-residue subset, APG-L achieves the best coverage F1 (61.61%) together with 99.71% conformer-level tetrahedral accuracy. CycPepFlow thus combines broad reference-ensemble coverage with high generated-set precision and stereochemical fidelity, with its six-mer performance highlighting the potential for extension to larger macrocycles.

1 Introduction

Cyclic peptides bridge small molecules and biologics (Driggers et al., 2008; Vinogradov et al., 2019; Muttenthaler et al., 2021). Their constrained scaffolds recognize extended protein surfaces (Villar et al., 2014; Jwad et al., 2020), while noncanonical amino acids (NCAAs) expand chemical space and can improve membrane permeability and oral absorption (White et al., 2011; Rezai et al., 2006; Nielsen et al., 2017). Computational design has yielded structurally ordered, membrane-permeable, and chemically diverse macrocycles (Hosseinzadeh et al., 2017; Bhardwaj et al., 2022; Salveson et al., 2024; Kamenik et al., 2018; Grambow et al., 2024b).

A closed ring couples local chemical geometry to long-range conformation (Kamenik et al., 2018; Grambow et al., 2024b). Protein-derived models capture global dependencies, but diverse NCAAs and cyclization chemistries remain challenging (Mao et al., 2025; Cao et al., 2025a). Cyclic positional encodings alone do not specify the closing linkage (Rettie et al., 2025a; Cao et al., 2025a), and NCAAs support for modifications in the Chemical Component Dictionary (CCD) does not ensure stereochemically valid coordinates (Abramson et al., 2024). Small-molecule conformer generators avoid fixed residue vocabularies, but their transfer to macrocycles requires explicit evaluation (Grambow et al., 2024b; Nikitin et al., 2025).

CycPepFlow learns Cartesian coordinate flows from graph-harmonic priors (Lipman et al., 2023; Hassan et al., 2024). SO(3)-equivariant local scalar–vector messages combine with gated global attention and optional graph-geodesic bias. Atomic chirality conditioning and signed-volume supervision explicitly target tetrahedral orientation. We evaluate CREMP/RINGER four- to six-residue macrocycles (Grambow et al., 2024a;b) using bidirectional coverage/matching and conformer-/center-level tetrahedral agreement, separating reference coverage from generated-set precision. Our contributions are:

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.

• **Topology- and stereochemistry-aware conformer generation**. We introduce a Cartesian flow model that generates cyclic-peptide conformer ensembles directly from molecular graphs. CycPepFlow combines typed local chemical interactions and molecule-wide attention with atomic chirality conditioning and signed-volume supervision to jointly learn coupled macrocyclic geometry and tetrahedral stereochemistry.

• **Attention and optimization for molecular flow modeling.** We incorporate Attention Residuals (AttnRes; Kimi Team et al., 2026) and gated attention (Qiu et al., 2025) from recent language-model research into cyclic-peptide conformer generation and evaluate their contributions through component ablations. We further find SOAP (Vyas et al., 2025) effective for training CycPepFlow under the evaluated configuration, extending prior evidence of its utility in interatomic-potential learning (Harari et al., 2026) to this cyclic-peptide flow model.

• **High-precision macrocyclic ensemble generation.** On the shared CREMP/RINGER benchmark, CycPepFlow variants achieve the highest overall coverage F1 and generated-set precision among the compared methods, including recent state-of-the-art learned conformer generators.

2 Related Work

2.1 Physics- and heuristic-based conformer generation

Classical conformer generation combines geometric constraints, conformational search, and energetic refinement. ETKDGv3 incorporates experimental torsional preferences and macrocycle-specific refinements into distance geometry (Riniker & Landrum, 2015; Wang et al., 2020). OMEGA, including its dedicated macrocycle mode, and Prime macrocycle sampling provide established alternatives for exploring candidate conformations (Hawkins et al., 2010; Sindhikara et al., 2017; Grambow et al., 2024b). However, ring closure and coupled backbone torsions complicate ensemble generation: ETKDGv3 and OMEGA Macrocycle showed limited ensemble precision in the reported benchmark (Grambow et al., 2024b). Molecular dynamics and enhanced sampling investigate solution ensembles, but transitions among distinct macrocyclic states can require prolonged simulations (Kamenik et al., 2018; Damjanovic et al., 2021). Ensemble accuracy also depends on the force field and solvent model, including appropriate parameterization of noncanonical and N-methylated residues (Kamenik et al., 2018; Damjanovic et al., 2021). CREST combines RMSD-biased metadynamics, geometry optimization, duplicate removal, and genetic recombination of internal-coordinate structures for low-energy conformer search (Pracht et al., 2024). In CREMP, CREST uses semiempirical GFN2-xTB calculations and implicit chloroform solvation to generate ensembles that agree well with a limited set of NMR structures in that solvent, albeit at substantial computational cost (Grambow et al., 2024a). CycPepFlow learns from CREMP conformer ensembles to generate conformers conditioned on molecular graphs.

2.2 Generative models for molecular conformations

Learned conformer generators differ in their molecular representations and geometric inductive biases. DMCG directly predicts atomic coordinates using a generative latent-variable model, whereas GeoMol assembles conformers from learned local geometry and torsions (Zhu et al., 2022; Ganea et al., 2021). GeoDiff generates Cartesian coordinates through equivariant diffusion, while torsional diffusion operates on rotatable-bond torsions while preserving local molecular geometry (Xu et al., 2022; Jing et al., 2022). More recent approaches explore a broader range of architectures and chemical conditioning strategies. ET-Flow applies equivariant flow matching to molecular coordinates (Hassan et al., 2024), while Molecular Conformer Fields (MCF) investigates scalable generation without explicit rotational equivariance (Wang et al., 2024). DiTMC separates coordinate processing from graph conditioning, compares non-equivariant and SO(3)-equivariant attention, and incorporates all-pair graph-geodesic information (Frank et al., 2025). LoQI focuses on low-energy, stereochemistry-aware generation using ChEMBL3D, encoding target stereochemistry through graph augmentation (Nikitin et al., 2025; Reidenbach et al., 2025). However, for macrocycles, ring closure couples torsions throughout the ring, making local geometric choices mutually dependent (Grambow et al., 2024b). Consequently, strong performance on small-molecule benchmarks does not necessarily translate to accurate cyclic peptide conformations (Section 5.2; Tables 4–6).

2.3 PROTEIN-DERIVED CYCLIC-PEPTIDE PREDICTION AND DESIGN

AlphaFold-like and related biomolecular models have been adapted for cyclic-peptide prediction and design. Methods based on AlphaFold2 (Jumper et al., 2021) include AfCycDesign (Rettie et al., 2025a), HighFold (Zhang et al., 2024), HighFold2 (Zhu et al., 2025), and HighFold-MeD (Cao et al., 2025b). AF3-based methods include HighFold3 (Cao et al., 2025a) and HighFold4 (Zhang et al., 2026). RoseTTAFold-based methods include NCPepFold (Mao et al., 2025) and RFpeptides (Rettie et al., 2025b). Boltz-based methods (Wohlwend et al., 2025; Passaro et al., 2025) include CyclicBoltz1 (Xie et al., 2025) and HighFold-MeD2 (Cao et al., 2026). Design approaches based on Protenix (ByteDance AML AI4Science Team et al., 2025) include PXDesign-d and PXDesign-h (Protenix Team et al., 2025), while Chai-based workflows include the Chai-1 branch of Protein Hunter (Cho et al., 2025). Despite these advances, NCAA support remains uneven: CCD-based workflows may require user-defined residue representations for modifications absent from the dictionary (Han et al., 2026). Moreover, input stereochemical specification does not guarantee its preservation in predicted coordinates (Cao et al., 2025b). These limitations motivate explicit evaluation of chemical coverage and stereochemical fidelity.

2.4 LEARNING CYCLIC-PEPTIDE CONFORMATIONAL ENSEMBLES

Learning from molecular dynamics has enabled ensemble prediction for cyclic pentapeptides and hexapeptides (Miao et al., 2021; Hui et al., 2023). RINGER directly addresses macrocyclic conformer generation through diffusion over redundant internal coordinates, cyclically invariant positional encoding, and constrained reconstruction (Grambow et al., 2024b). In the Cartesian setting, AGDIFF provides an attention-enhanced molecular diffusion architecture; Wu and Zou retrain it on CREMP and add post-generation stereochemical correction for cyclic-peptide prediction (Vieira Wyzykowski et al., 2025; Wu & Zou, 2026). Vilya-1 is an all-atom diffusion approach to chemically diverse macrocycle conformer generation, with reported improvements in recovering experimental structures and extensions to confidence estimation, property prediction, and design (Vilya Research et al., 2026b). Vilya-2 advances this framework with improved conformer generation and receptor-conditioned peptide-interface modeling, including confidence-ranked pose prediction (Vilya Research et al., 2026a). Vilya-1 and Vilya-2 broaden the scope of macrocycle modeling, but were not available to us for independent retraining and evaluation.

3 CYCPEPFLOW

3.1 TASK AND CONDITIONAL COORDINATE FLOW

For a fixed connected molecular graph G with N explicit atoms, including hydrogens, we generate arithmetic-mean-centered coordinates $X \in \mathbb{R}^{N\times 3}$ conditioned on atom identity, connectivity, formal charge, and tetrahedral stereochemistry. Figure 1a illustrates the graph-conditioned flow from a prior sample to a generated conformer. The vector-field network in Figure 1b predicts coordinate velocities that transport a graph-harmonic Gaussian prior, whose covariance on the centered subspace is the pseudoinverse of the covalent graph Laplacian (Jing et al., 2023; Stark et al., 2024; Hassan et al., 2024; Appendix A.1). Hydrogens participate throughout generation and are hidden only in the molecular renderings for visual clarity.

During training, the prior coordinates X_0 are aligned to the observed conformer X_1 by a proper-rotation Kabsch transformation (Kabsch, 1976). For uniformly sampled flow time $t\sim \mathcal{U}\left(10^{-4}, 0.9999\right)$, centered Gaussian noise ϵ, and noise amplitude $\sigma = 0.1$, the intermediate coordinates X_t and their samplewise velocity u_t are

$$X_t = (1-t)X_0 + tX_1 + \sigma\sqrt{t(1-t)}\epsilon, \tag{1}$$

$$u_t = X_1 - X_0 + \frac{\sigma(1-2t)}{2\sqrt{t(1-t)}}\epsilon. \tag{2}$$

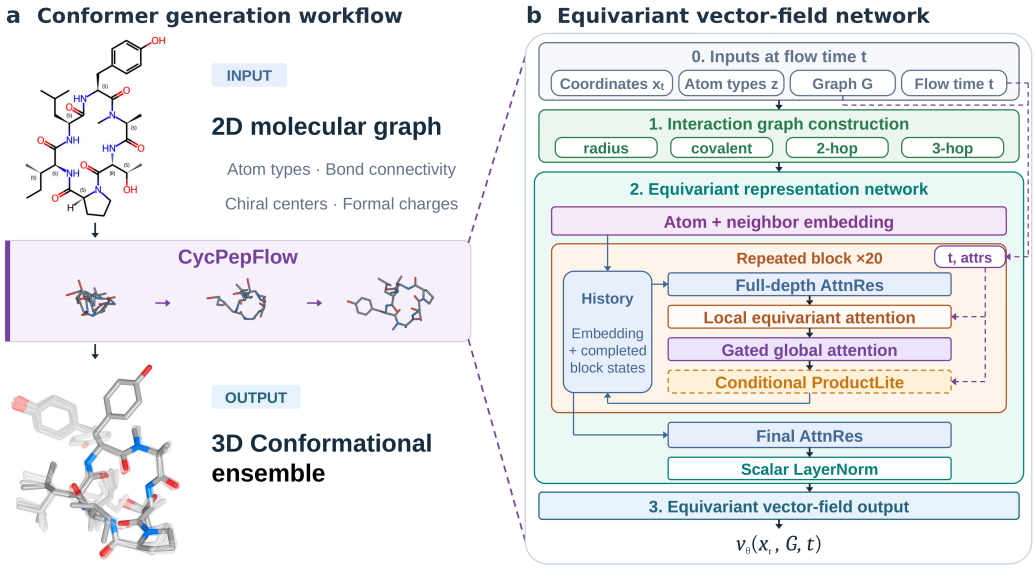


Figure 1: CycPepFlow workflow and vector-field network. (a) Schematic generation trajectories for the illustrated molecule, with coordinate snapshots at $t = 0$, 0.5, and 1. Hydrogens are included in the model but hidden in the renderings for clarity. (b) The network predicts the coordinate velocity at the current time. It constructs the interaction graph and applies 20 independently parameterized blocks. Attention Residuals (AttnRes) mixes the initial embedding and completed block states, including a final mixture at readout. Conditional ProductLite acts in blocks 5, 10, 15, and 20. Atom attributes encode formal charge and chirality, and time is supplied as a raw scalar. Global attention updates scalar features, and the vector readout returns centered atomic velocities.

For a local minibatch of B molecules, the velocity-regression loss $\mathcal{L}_{vel}$ is

$$\mathcal{L}_{vel} = \frac{1}{B}\sum_{g=1}^{B}\frac{1}{N_g}\sum_{i=1}^{N_g} \| \, v_{\theta,gi}\left(X_{t_g}, G_g, t_g\right) - u_{t_g,i} \, \|_2 \, . \tag{3}$$

Here g indexes molecules, i indexes atoms, and N_g is molecule g's atom count. With learned parameters θ, $v_{\theta,gi}$ is the predicted velocity of its atom i given coordinates X_{t_g}, molecular graph G_g, and sampled time t_g; $u_{t_g,i}$ is the corresponding target velocity. The executed loss averages unsquared Euclidean errors over atoms and then molecules (Lipman et al., 2023; Hassan et al., 2024). At inference, integrating the learned vector field transforms a prior sample into a conformer, as illustrated schematically in Figure 1a.

3.2 Topology and Pair Geometry

Figure 2 summarizes the node features and typed interaction edges. Atomic number is embedded, while the RDKit chirality-category and formal-charge-category indices are transformed by a node MLP. For a tetrahedral atom, RDKit's clockwise (CW) or counterclockwise (CCW) tag encodes the orientation relative to the molecule's stored neighbor order. Each vector-field evaluation combines exact one-, two-, and three-hop covalent pairs with current-coordinate radius-only pairs, with topology taking precedence. Radius neighbors are queried within 10 Å with a 32-neighbor cap before the union. The resulting graph and its geometry are reused by all 20 network layers at the current coordinates. After an Euler update changes the coordinates, the next network call recomputes the radius neighbors and geometric edge features; the covalent connectivity and its hop relations remain fixed.

For an incoming edge $j \to i$, where j and i index source and receiving atoms with current positions $X_{t,j}$ and $X_{t,i}$, the implementation uses

$$\rho_{ij} = \| \, X_{t,j} - X_{t,i} \, \|_2^2, \qquad \delta_{ij} = \frac{X_{t,j} - X_{t,i}}{\max\left(\rho_{ij}, 10^{-2}\right)}. \tag{4}$$

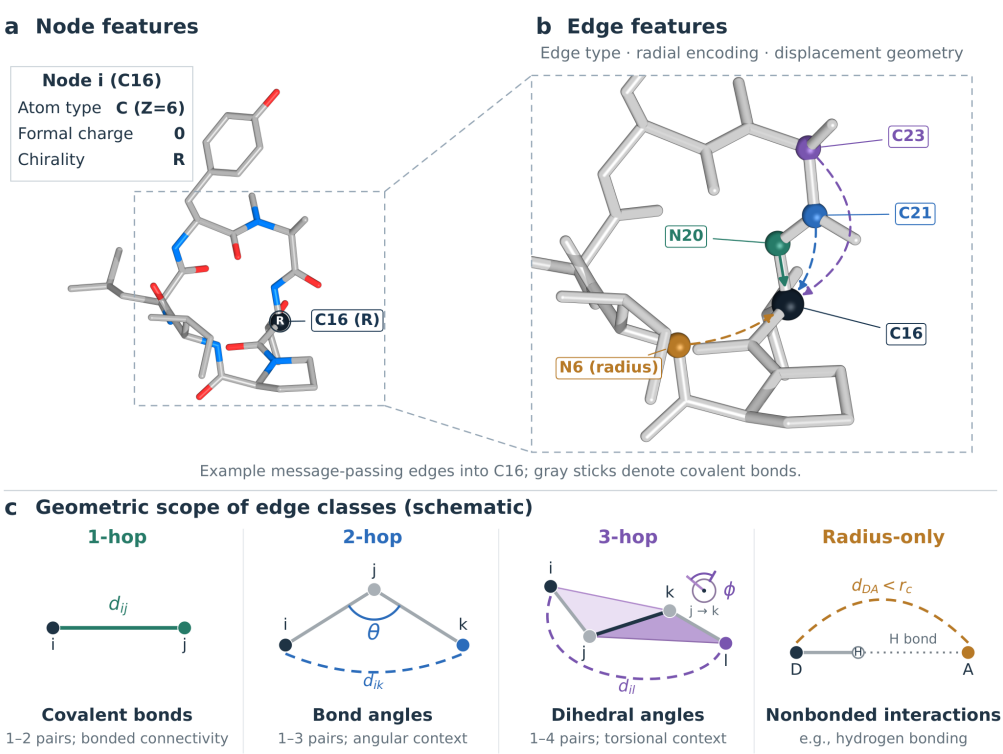


Figure 2: Node attributes and typed interaction edges. (a) Atomic identity, formal charge, and tetrahedral stereochemistry are illustrated for a representative atom. (b) Directed computational interactions distinguish covalent one-hop, exact two-hop, exact three-hop, and radius-only pairs, together with radial and displacement features. Gray sticks show chemical bonds; colored arrows show message-passing relations. (c) Bonded, angular, torsional, and nonbonded cartoons illustrate the chemical motivations for these edge classes. Hydrogens are hidden for visual clarity but are retained in the molecular graph and all model computations.

Here ρ_{ij} is squared separation and δ_{ij} is the scaled displacement. The 64 trainable radial functions receive ρ_{ij} alongside a relation code. Their scalar cosine envelope ends at physical distance $\sqrt{10}$ Å, distinct from the 10 Å radius query. Vector messages omit this envelope; Appendix A gives the exact construction.

3.3 Local–Global Equivariant Representation

Each atom i carries invariant scalar features $h_i \in \mathbb{R}^d$ and vector features $V_i \in \mathbb{R}^{3\times d}$, where d is the channel width. Local blocks condition on time t and node attributes and follow the TorchMD-Net scalar–vector pattern (Thölke & De Fabritiis, 2022), schematically

$$m_{ij}^V = V_j \odot b_{ij}^{(1)} + \delta_{ij} \otimes b_{ij}^{(2)} + (\delta_{ij} \times V_j) \odot b_{ij}^{(3)}. \tag{5}$$

Here m_{ij}^V is the vector message from atom j to i, and $b_{ij}^{(1)}, b_{ij}^{(2)}, b_{ij}^{(3)}$ are learned channelwise coefficients. The learned coefficients depend on conditioned atom features and edge geometry. Only the scalar messages use a separate local-attention multiplier. The cross-product term preserves SO(3) equivariance, allowing the vector pathway to retain orientation-sensitive geometric information. Figure 6 details the local computation.

After each local update, gated global attention communicates across all atoms in the molecule (Figure 7), using headwise output gating adapted from Qiu et al. (2025). Full-depth Attention Residuals (Kimi Team et al., 2026) mixes the initial embedding and completed block states with

shared invariant weights for the scalar and vector streams, including a final readout mixture (Figure 5). Conditional ProductLite adds a lightweight MACE-inspired residual (Batatia et al., 2022; Figure 8). Ablations in Figure 9 assess these design choices; detailed results and control limitations are given in Appendix C.

Optional all-pair graph-geodesic bias. Following shortest-path bias in graph attention (Ying et al., 2021) and molecular graph-geodesic conditioning in DiTMC (Frank et al., 2025), APG augments each global-attention head:

$$\ell_{ij}^{(m)} = \frac{\left(q_i^{(m)}\right)^{\mathsf{T}} k_j^{(m)}}{\sqrt{d_h}} + B_{ij}^{(m)}. \tag{6}$$

The bias $B_{ij}^{(m)} = b_{\ell,m}\left[\kappa(i,j)\right]$ uses layer ℓ's lookup table $b_{\ell,m}$ for head m, indexed by the binned covalent shortest-path length $\kappa(i,j)$; it is zero without APG. Connected paths are clipped to 32, with bin 33 for disconnection/padding. Appendix A.5 gives the full attention specification.

3.4 Stereochemical supervision

Inspired by auxiliary chirality supervision in Vilya-1 (Vilya Research et al., 2026b), we use a normalized signed-volume objective. For each usable source-specified tetrahedral center i, the first three sorted explicit-neighbor indices define the normalized scalar triple product $\bar{\nu}_i(X)$ and reference sign s_i. For the extrapolated endpoint $\widehat{X}_1 = X_t + (1-t)v_\theta\left(X_t, G, t\right)$,

$$\mathcal{L} = \mathcal{L}_{vel} + 0.5\mathcal{L}_{chiral}, \qquad \mathcal{L}_{chiral} = \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} softplus\left[-s_i\bar{\nu}_i\left(\widehat{X}_1\right)\right]. \tag{7}$$

Here $\mathcal{C}$ is the set of valid tetrahedral centers pooled over the local minibatch; $|\mathcal{C}|$ is its size. The chirality term averages over valid tetrahedral centers in each local minibatch, returns zero when no valid center is present, and is active from the first training step. Appendix A.7 specifies the signed-volume calculation and mask.

4 Evaluation

4.1 Dataset, training, and comparator scope

We follow the RINGER/CREMP data-splitting protocol (Grambow et al., 2024a;b), with 1,000 held-out test molecules comprising 533 four-mers, 347 five-mers, and 120 six-mers (Appendix E.1). The shared data protocol retains up to the 30 lowest-energy conformers per training or validation molecule, and uses all 877,898 test reference conformers. The shared generation protocol requests $2K$ conformers for a molecule with K reference conformers. Model-specific output protocols and literature-baseline provenance are detailed in Appendix E.

All four CycPepFlow variants are trained from scratch for 200 epochs with the same random seed, the SOAP optimizer (Vyas et al., 2025), a learning rate of 8×10^{-4}, and an effective global batch size of 64. Checkpoints are selected by validation loss. Training runs on NVIDIA GeForce RTX 4090 D GPUs. Inference uses 50 deterministic Euler steps from $t = 0$ to $t = 1$, with BF16 network autocast for acceleration. Appendix A.8 provides the complete training and inference settings.

To compare with state-of-the-art learned conformer generators, we retrain LoQI, ET-Flow, MCF-S, and the DiTMC aPE/rPE B/L variants on CREMP. The "Macro" suffix distinguishes these CREMP-retrained models from their original published versions. We use model-specific data adapters and perform training, checkpoint selection, conformer generation, and scoring under the shared molecular split and top-30 training-conformer policy. Each model family's architecture and training procedure are retained, subject to the documented CREMP adaptations; Appendix E lists the actual training budgets and selected checkpoints. In particular, MCF-S follows a step-based schedule rather than CycPepFlow's 200-epoch schedule and uses a validation-selected exponential moving average (EMA) checkpoint at 300,000 steps. The six dagger-marked rows in Figure 3 report results from the published RINGER comparison; these baselines were not retrained in this study.

4.2 BIDIRECTIONAL ENSEMBLE AND STEREOCHEMICAL METRICS

We evaluate ensemble quality in both directions to distinguish reference coverage from generated-set precision. For molecule m, let R_m and S_m denote the reference and generated conformer sets, respectively. For reference conformer r and generated conformer g, D_{rg} is their optimally aligned heavy-atom root-mean-square deviation (RMSD). The nearest-conformer distances are $d_r = \min_{g \in S_m} D_{rg}$ and $d_g = \min_{r \in R_m} D_{rg}$. Recall (R) measures reference-ensemble coverage, while precision (P) measures generated-ensemble agreement. With coverage threshold $\delta = 0.75$ Å, the recall and precision metrics are

$$COV - R_m = \frac{100}{|R_m|} \sum_r \mathbf{1}\left[d_r < \delta\right], \qquad MAT - R_m = \frac{1}{|R_m|} \sum_r d_r, \tag{8}$$

$$COV - P_m = \frac{100}{|S_m|} \sum_g \mathbf{1}\left[d_g < \delta\right], \qquad MAT - P_m = \frac{1}{|S_m|} \sum_g d_g. \tag{9}$$

Here $|R_m|$ and $|S_m|$ are the reference and generated conformer counts, respectively, and $\mathbf{1}[\cdot]$ is the indicator function. COV is a percentage, with higher values preferred, whereas MAT is a distance in Å, with lower values preferred. Because an inherent trade-off exists between recall and precision—where maximizing one often comes at the expense of the other—we also report their harmonic mean (F1) to capture the overall performance.

We evaluate stereochemical fidelity at input-specified tetrahedral centers. Conformer-level accuracy is the percentage of generated conformers whose stereocenters all match their reference configurations, excluding conformers with no stereocenters. Stereocenter-level accuracy is the percentage of matching centers pooled across generated conformers. Near-planar generated centers are counted as incorrect. Appendix D specifies the reference configurations and the conformer and stereocenter counts used to calculate these percentages.

5 RESULTS

5.1 CYCPEPFLOW ACHIEVES THE BEST OVERALL ENSEMBLE QUALITY WITH A FAVORABLE PRECISION–RECALL BALANCE

Figure 3 compares conformational ensemble quality across 1,000 CREMP test molecules, with complete numerical results provided in Table 3. DMCG-Macro achieves the highest reference coverage (COV-R: 78.30%), whereas DiTMC aPE-L-Macro attains the lowest reference matching distance (MAT-R: 0.495 Å). In contrast, CycPepFlow APG-L achieves the highest generated-set coverage (COV-P: 79.78%) and the lowest generated-set matching distance (MAT-P: 0.419 Å), indicating the strongest generated-ensemble precision among the evaluated methods. By combining competitive recall with high precision, the CycPepFlow variants achieve the best overall ensemble scores: CycPepFlow-L yields the highest COV-F1 (76.13%), while CycPepFlow APG-L yields the lowest MAT-F1 (0.477 Å).

APG could shift the precision–recall balance toward precision. Compared with CycPepFlow-L, CycPepFlow APG-L increases COV-P by 1.57 percentage points and reduces MAT-P from 0.441 to 0.419 Å, with a modest reduction in reference coverage. Nevertheless, COV-F1 remains nearly unchanged (76.05% versus 76.13%), while MAT-F1 improves from 0.485 to 0.477 Å. High precision means that a smaller fraction of generated conformers falls outside the reference-matching threshold, yielding fewer reference-inconsistent candidates and reducing the burden of downstream conformer screening. APG therefore offers a useful precision-oriented alternative.

5.2 CYCPEPFLOW MAINTAINS HIGH STEREOCHEMICAL ACCURACY AND STRONGER SIX-MER PERFORMANCE

Figure 4 compares ensemble quality and stereochemical accuracy across the four-, five-, and six-mer subsets of the CREMP test set. Both CycPepFlow variants maintain conformer-level stereochemical accuracy above 99.4% and stereocenter-level accuracy of at least 99.9% across all three lengths. Their

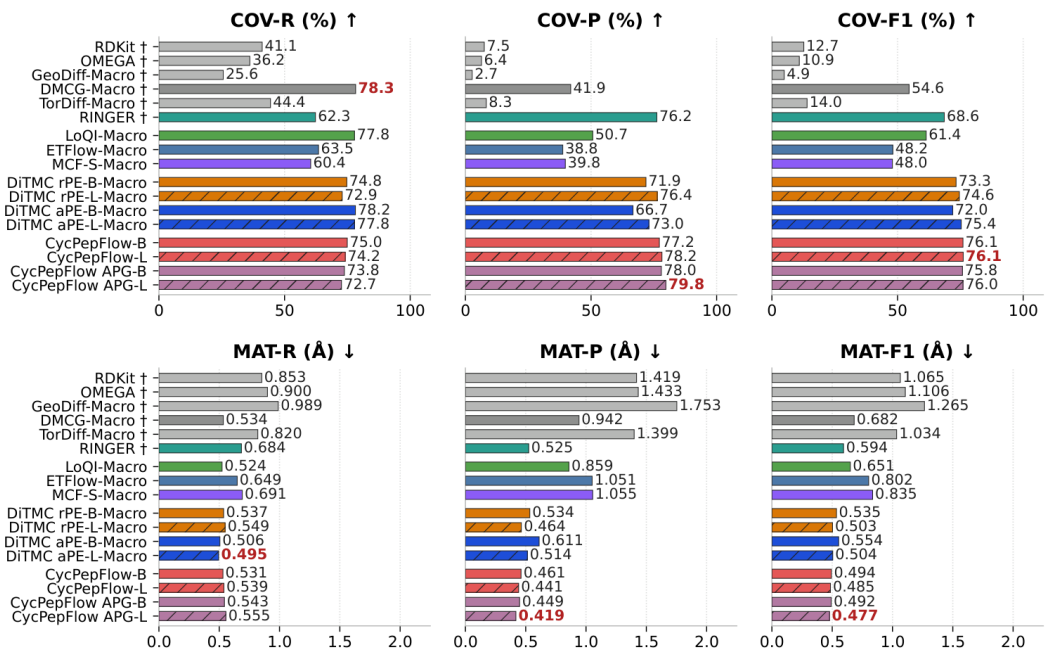


Figure 3: Conformer ensemble generation benchmark on CREMP. Coverage/matching summaries for 1,000 test molecules. Dagger-marked rows are results reported in the RINGER benchmark. Other rows are models trained and evaluated on CREMP in this study. Direct row labels identify methods; solid/hatched bars distinguish B/L within DiTMC and CycPepFlow. Red/bold values mark extrema before rounding. Table 3 provides all values.

advantage is particularly pronounced for six-mers, where CycPepFlow-L and CycPepFlow APG-L achieve conformer-level accuracies of 99.44% and 99.71%, respectively, compared with 92.86% for LoQI, 91.30% for DiTMC rPE-L, and 93.14% for DiTMC aPE-L. ETFlow and MCF-S achieve substantially lower accuracies of 25.49% and 24.53%, respectively. At the individual-stereocenter level, CycPepFlow-L and CycPepFlow APG-L achieve six-mer accuracies of 99.91% and 99.95%, exceeding the strongest non-CycPepFlow baseline, LoQI, at 98.89%. Thus, CycPepFlow reliably preserves the specified tetrahedral stereochemistry not only at individual centers but also across complete conformers, especially for the largest rings evaluated.

CycPepFlow also achieves the strongest balanced ensemble performance on the six-mer subset. CycPepFlow APG-L attains the highest COV-F1 (61.61%), outperforming CycPepFlow-L (58.66%) and DiTMC rPE-L (57.71%), the strongest non-CycPepFlow baseline on this metric. Relative to CycPepFlow-L, incorporating APG increases COV-P by 4.35 percentage points and COV-F1 by 2.95 percentage points, reduces MAT-P by 0.066 Å, and preserves essentially the same COV-R (71.78% versus 71.75%). APG-L also achieves the lowest six-mer MAT-F1 (0.733 Å). These gains suggest that APG helps maintain ensemble quality as ring size increases within the evaluated range. Together with the consistently high stereochemical accuracy, these results highlight CycPepFlow's effectiveness across four- to six-mers and support the potential of APG to facilitate its extension to larger cyclic peptides.

6 Limitations and Future Directions

Several limitations of CycPepFlow also motivate future development. First, reflection symmetry is not explicitly enforced: because the local cross-product construction mixes polar- and axial-vector contributions, reflection equivariance is not guaranteed. Consequently, two molecular configurations related by reflection are expected to yield correspondingly reflected predictions in principle, but this relationship is not enforced by the architecture, although we observe that their generated structures are highly similar in practice. Second, while the strong six-mer results highlight the potential

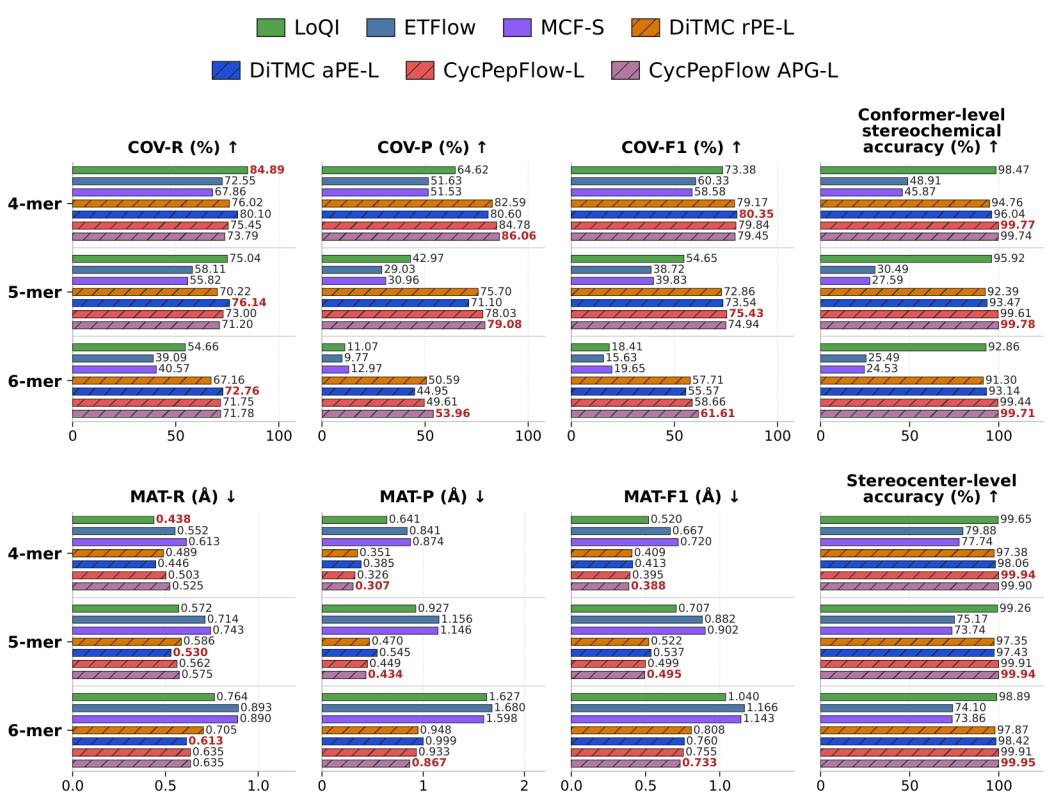


Figure 4: Length-stratified ensemble quality and stereochemical accuracy on CREMP. Seven models are evaluated on subsets containing 533 four-mers, 347 five-mers, and 120 six-mers. Colors and hatch patterns are consistent with Figure 3. Stereochemical accuracy measures agreement with the tetrahedral stereochemical assignments in the source molecules: conformer-level accuracy requires agreement at all evaluated centers within a conformer, whereas stereocenter-level accuracy pools agreement over individual centers. Bold red values indicate the best result for each metric within each length group. Tables 4–6 provide the complete numerical results.

of CycPepFlow to scale to larger cyclic peptides, evaluation beyond six residues, across broader noncanonical chemistry, alternative cyclization linkages, and protein-bound conformations will be needed to establish its generalizability. Computational scaling also remains a consideration, as dense global attention scales quadratically with atom count and the current APG shortest-path preprocessing scales cubically; topology caching and more efficient graph processing could facilitate applications to larger systems. Finally, the present study focuses on conformer generation for fixed molecular graphs. Extending CycPepFlow toward property-aware and target-conditioned cyclic-peptide design, with appropriate conditioning mechanisms and task-specific validation, represents an important direction for future work.

7 Conclusion

CycPepFlow integrates local geometric interactions, molecule-wide attention, and explicit stereochemical learning within a Cartesian flow framework for cyclic-peptide ensemble generation. On CREMP, its variants achieve the highest overall coverage F1 and strongest generated-set precision among the compared methods while maintaining high tetrahedral accuracy. The optional graph-geodesic attention bias offers a precision-oriented alternative and improves balanced ensemble performance on six-residue peptides. Future work will evaluate generalization to larger rings and explore property-aware and target-conditioned cyclic-peptide design.

AI USE STATEMENT

Generative AI tools were used to discuss and refine research methodology, assist software implementation and editing, retrieve and organize literature, interpret and tabulate benchmark results, prepare scientific figures, and revise the manuscript. All reported numerical results were obtained from recorded computations or cited literature sources. The authors reviewed all AI-assisted content and checked the text, equations, citations, and tables against the underlying sources. The authors take full responsibility for the accuracy, originality, and integrity of the final manuscript and all associated AI-assisted materials.

ETHICS STATEMENT

This study uses the existing, computationally generated CREMP dataset and does not involve human participants or personal data. CycPepFlow is intended to support cyclic-peptide conformational research, not to establish biological activity or clinical suitability. Agreement with computational reference conformations does not establish target binding, membrane permeability, therapeutic efficacy, or safety. Inaccurate or incomplete ensembles may mislead downstream analyses and molecular design; the benchmark scope and limitations are therefore reported explicitly. Any downstream experimental or design application requires independent validation and appropriate scientific, ethical, and safety review, including assessment of potential dual-use risks.

REPRODUCIBILITY STATEMENT

Sections 3–4 and Appendices A–B specify the model, flow construction, graph representation, chirality objective, and training and inference settings. Appendices C–F provide ablation details, metric definitions and aggregation rules, data-split and comparator provenance, and complete numerical results for Figures 3–4 and 9, distinguishing literature baselines from models evaluated in this study. Machine-readable results and figure PDFs are included in the accompanying source package. The code repository provides inference and evaluation code, pretrained checkpoints, data-split manifests, and environment setup instructions. As an internal reproducibility check, another lab member reran the packaged inference and evaluation workflow in a clean software environment, reproducing the reported held-out test metrics for all four CycPepFlow variants within the documented tolerances.

REFERENCES

Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J. Ballard, Joshua Bambrick, Sebastian W. Bodenstein, David A. Evans, Chia-Chun Hung, Michael O'Neill, David Reiman, Kathryn Tunyasuvunakool, Zachary Wu, Akvilė Žemgulytė, Eirini Arvaniti, Charles Beattie, Ottavia Bertolli, Alex Bridgland, Alexey Cherepanov, Miles Congreve, Alexander I. Cowen-Rivers, Andrew Cowie, Michael Figurnov, Fabian B. Fuchs, Hannah Gladman, Rishub Jain, Yousuf A. Khan, Caroline M. R. Low, Kuba Perlin, Anna Potapenko, Pascal Savy, Sukhdeep Singh, Adrian Stecula, Ashok Thillaisundaram, Catherine Tong, Sergei Yakneen, Ellen D. Zhong, Michal Zielinski, Augustin Žídek, Victor Bapst, Pushmeet Kohli, Max Jaderberg, Demis Hassabis, and John M. Jumper. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. doi: 10.1038/s41586-024-07487-w. URL https://doi.org/10.1038/s41586-024-07487-w.

Ilyes Batatia, Dávid Péter Kovács, Gregor N. C. Simm, Christoph Ortner, and Gábor Csányi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. In *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436, 2022. doi: 10.52202/068431-0830. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/4a36c3c51af11ed9f34615b81edb5bbc-Abstract.html. arXiv:2206.07697.

Gaurav Bhardwaj, Jacob O'Connor, Stephen Rettie, Yen-Hua Huang, Theresa A. Ramelot, Vikram Khipple Mulligan, Gizem Gokce Alpkilic, Jonathan Palmer, Asim K. Bera, Matthew J. Bick, Maddalena Di Piazza, Xinting Li, Parisa Hosseinzadeh, Timothy W. Craven, Roberto Tejero, Anna Lauko, Ryan Choi, Calina Glynn, Linlin Dong, Robert Griffin, Wesley C. van Voorhis, Jose Rodriguez, Lance Stewart, Gaetano T. Montelione, David Craik, and David Baker. Accurate de novo design of membrane-traversing macrocycles. *Cell*, 185(19):3520–3532.e26, 2022. doi: 10.1016/j.cell.2022.07.019. URL https://doi.org/10.1016/j.cell.2022.07.019.

ByteDance AML AI4Science Team, Xinshi Chen, Yuxuan Zhang, Chan Lu, Wenzhi Ma, Jiaqi Guan, Chengyue Gong, Jincai Yang, Hanyu Zhang, Ke Zhang, Shenghao Wu, Kuangqi Zhou, Yanping Yang, Zhenyu Liu, Lan Wang, Bo Shi, Shaochen Shi, and Wenzhi Xiao. Protenix - Advancing Structure Prediction Through a Comprehensive AlphaFold3 Reproduction. *bioRxiv*, 2025. doi: 10.1101/2025.01.08.631967. URL https://www.biorxiv.org/content/10.1101/2025.01.08.631967v1. Version 1, posted 2025-01-11.

Sen Cao, Cheng Zhu, Qingyi Mao, Jingjing Guo, Ning Zhu, and Hongliang Duan. Accurate structure prediction of cyclic peptides containing unnatural amino acids using HighFold3. *Briefings in Bioinformatics*, 26(5):bbaf488, 2025a. doi: 10.1093/bib/bbaf488. URL https://doi.org/10.1093/bib/bbaf488.

Zhigang Cao, Sen Cao, Linghong Wang, Zhiguo Wang, Qingyi Mao, Jingjing Guo, and Hongliang Duan. HighFold-MeD: a Rosetta distillation model to accelerate structure prediction of cyclic peptides with backbone N-methylation and d-amino acids. *Journal of Cheminformatics*, 17(1):167, 2025b. doi: 10.1186/s13321-025-01111-3. URL https://doi.org/10.1186/s13321-025-01111-3.

Zhigang Cao, Wentong Wang, Sen Cao, Chengyun Zhang, Jingjing Guo, and Hongliang Duan. HighFold-MeD2: An Enhanced Boltz-2 Model for Accurate Structure Prediction of N-Methylated and d-Amino Acid Cyclic Peptides. *Journal of Chemical Information and Modeling*, 66(10):6057–6066, 2026. doi: 10.1021/acs.jcim.6c00158. URL https://doi.org/10.1021/acs.jcim.6c00158.

Yehlin Cho, Griffin Rangel, Gaurav Bhardwaj, and Sergey Ovchinnikov. Protein Hunter: exploiting structure hallucination within diffusion for protein design. *bioRxiv*, 2025. doi: 10.1101/2025.10.10.681530. URL https://www.biorxiv.org/content/10.1101/2025.10.10.681530v2. Preprint, version 2, posted 2025-10-12.

Jovan Damjanovic, Jiayuan Miao, He Huang, and Yu-Shan Lin. Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations. *Chemical Reviews*, 121(4):2292–2324,

2021. doi: 10.1021/acs.chemrev.0c01087. URL https://doi.org/10.1021/acs.chemrev.0c01087.

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. doi: 10.1038/nrd2590. URL https://doi.org/10.1038/nrd2590.

J. Thorben Frank, Winfried Ripken, Gregor Lied, Klaus-Robert Müller, Oliver Unke, and Stefan Chmiela. Sampling 3D Molecular Conformers with Diffusion Transformers. In *Advances in Neural Information Processing Systems*, volume 38, pp. 168881–168931, 2025. doi: 10.52202/085713-5626. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/f6fbaf131fa71b036d7b8c49379f3af4-Paper-Conference.pdf.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. GeoMol: Torsional Geometric Generation of Molecular 3D Conformer Ensembles. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 13757–13769. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/725215ed82ab6306919b485b81ff9615-Paper.pdf.

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, 2024a. doi: 10.1038/s41597-024-03698-y. URL https://doi.org/10.1038/s41597-024-03698-y.

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:2305.19800v2*, 2024b. doi: 10.48550/arXiv.2305.19800. URL https://arxiv.org/abs/2305.19800v2. Version 2, revised August 13, 2024; first posted May 30, 2023.

Yanchao Han, Jianfeng Mei, Gaoshuai Li, Enkang Dai, Hanlei Lu, Chengyun Zhang, Yanlu Zhang, Chenshui Lin, Chuanlong Zeng, Hongliang Duan, and Xudong Wang. HighRes_Builder: improved access and modeling of noncanonical residues for protein structure prediction. *Briefings in Bioinformatics*, 27(3):bbag272, 2026. doi: 10.1093/bib/bbag272. URL https://doi.org/10.1093/bib/bbag272.

Gil Harari, Yoel Zimmermann, Ola Tangen Kulseng, Laura Zichi, Chuin Wei Tan, Marc L. Descoteaux, and Boris Kozinsky. Beyond Adam: SOAP and Muon for Faster, Label-Efficient Training of Machine Learning Interatomic Potentials. *arXiv preprint arXiv:2607.02499v1*, 2026. doi: 10.48550/arXiv.2607.02499. URL https://arxiv.org/abs/2607.02499v1. Version 1, posted July 2, 2026.

Majdi Hassan, Nikhil Shenoy, Jungyoon Lee, Hannes Stärk, Stephan Thaler, and Dominique Beaini. ET-Flow: Equivariant Flow-Matching for Molecular Conformer Generation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 128798–128824, 2024. doi: 10.52202/079017-4091. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/e8bd617e7dd0394ceadf37b4a7773179-Paper-Conference.pdf.

Paul C. D. Hawkins, A. Geoffrey Skillman, Gregory L. Warren, Benjamin A. Ellingson, and Matthew T. Stahl. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *Journal of Chemical Information and Modeling*, 50(4):572–584, 2010. doi: 10.1021/ci100031x. URL https://doi.org/10.1021/ci100031x.

Parisa Hosseinzadeh, Gaurav Bhardwaj, Vikram Khipple Mulligan, Matthew D. Shortridge, Timothy W. Craven, Fátima Pardo-Avila, Stephen A. Rettie, David E. Kim, Daniel-Adriano Silva, Yehia M. Ibrahim, Ian K. Webb, John R. Cort, Joshua N. Adkins, Gabriele Varani, and David Baker. Comprehensive computational design of ordered peptide macrocycles. *Science*, 358(6369):1461–1466, 2017. doi: 10.1126/science.aap7577. URL https://doi.org/10.1126/science.aap7577.

Tiffani Hui, Marc L. Descoteaux, Jiayuan Miao, and Yu-Shan Lin. Training Neural Network Models Using Molecular Dynamics Simulation Results to Efficiently Predict Cyclic Hexapeptide Structural Ensembles. *Journal of Chemical Theory and Computation*, 19(14):4757–4769, 2023. doi: 10.1021/acs.jctc.3c00154. URL https://doi.org/10.1021/acs.jctc.3c00154.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional Diffusion for Molecular Conformer Generation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24240–24253, 2022. URL https://arxiv.org/abs/2206.01729v2. arXiv:2206.01729.

Bowen Jing, Ezra Erives, Peter Pao-Huang, Gabriele Corso, Bonnie Berger, and Tommi Jaakkola. EigenFold: Generative protein structure prediction with diffusion models. *arXiv preprint arXiv:2304.02198*, 2023. doi: 10.48550/arXiv.2304.02198. URL https://arxiv.org/abs/2304.02198v1. Version 1; presented at the ICLR 2023 Machine Learning for Drug Discovery workshop.

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

Rasha Jwad, Daniel Weissberger, and Luke Hunter. Strategies for Fine-Tuning the Conformations of Cyclic Peptides. *Chemical Reviews*, 120(17):9743–9789, 2020. doi: 10.1021/acs.chemrev.0c00013. URL https://doi.org/10.1021/acs.chemrev.0c00013.

W. Kabsch. A Solution for the Best Rotation to Relate Two Sets of Vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. doi: 10.1107/S0567739476001873. URL https://doi.org/10.1107/S0567739476001873.

Anna S. Kamenik, Uta Lessel, Julian E. Fuchs, Thomas Fox, and Klaus R. Liedl. Peptidic Macrocycles - Conformational Sampling and Thermodynamic Characterization. *Journal of Chemical Information and Modeling*, 58(5):982–992, 2018. doi: 10.1021/acs.jcim.8b00097. URL https://doi.org/10.1021/acs.jcim.8b00097.

Kimi Team, Guangyu Chen, Yu Zhang, Jianlin Su, Weixin Xu, Siyuan Pan, Yaoyu Wang, Yucheng Wang, Guanduo Chen, Bohong Yin, Yutian Chen, Junjie Yan, Ming Wei, Y. Zhang, Fanqing Meng, Chao Hong, Xiaotong Xie, Shaowei Liu, Enzhe Lu, Yunpeng Tai, Yanru Chen, Xin Men, Haiqing Guo, Y. Charles, Haoyu Lu, Lin Sui, Jinguo Zhu, Zaida Zhou, Weiran He, Weixiao Huang, Xinran Xu, Yuzhi Wang, Guokun Lai, Yulun Du, Yuxin Wu, Zhilin Yang, and Xinyu Zhou. Attention Residuals. arXiv:2603.15031v1, 2026. URL https://arxiv.org/abs/2603.15031v1.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations*, 2023. URL https://iclr.cc/virtual/2023/poster/11309. arXiv:2210.02747.

Qingyi Mao, Tianfeng Shang, Wen Xu, Silong Zhai, Chengyun Zhang, Jingjing Guo, An Su, Chengxi Li, and Hongliang Duan. NCPepFold: Accurate Prediction of Noncanonical Cyclic Peptide Structures via Cyclization Optimization with Multigranular Representation. *Journal of Chemical Theory and Computation*, 21(9):4979–4991, 2025. doi: 10.1021/acs.jctc.5c00139. URL https://doi.org/10.1021/acs.jctc.5c00139.

Jiayuan Miao, Marc L. Descoteaux, and Yu-Shan Lin. Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chemical Science*, 12(44):14927–14936, 2021. doi: 10.1039/d1sc05562c. URL https://doi.org/10.1039/d1sc05562c.

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

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. doi: 10.1021/acs.chemrev.6b00838. URL https://doi.org/10.1021/acs.chemrev.6b00838.

Filipp Nikitin, Dylan M. Anstine, Roman Zubatyuk, Saee Gopal Paliwal, and Olexandr Isayev. Scalable Low-Energy Molecular Conformer Generation with Quantum Mechanical Accuracy. *ChemRxiv*, 2025. doi: 10.26434/chemrxiv-2025-k4h7v. URL https://client.prod.orp.cambridge.org/engage/chemrxiv/article-details/689f756ea94eede154ec6217. Preprint, version 1, posted August 20, 2025.

Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*, 2025. doi: 10.1101/2025.06.14.659707. URL https://www.biorxiv.org/content/10.1101/2025.06.14.659707v1. Version 1, posted 2025-06-18.

Philipp Pracht, Stefan Grimme, Christoph Bannwarth, Fabian Bohle, Sebastian Ehlert, Gereon Feldmann, Johannes Gorges, Marcel Müller, Tim Neudecker, Christoph Plett, Sebastian Spicher, Pit Steinbach, Patryk A. Wesołowski, and Felix Zeller. CREST—A program for the exploration of low-energy molecular chemical space. *The Journal of Chemical Physics*, 160(11):114110, 2024. doi: 10.1063/5.0197592. URL https://doi.org/10.1063/5.0197592.

Protenix Team, Milong Ren, Jinyuan Sun, Jiaqi Guan, Cong Liu, Chengyue Gong, Yuzhe Wang, Lan Wang, Qixu Cai, Wenzhi Ma, Yuxuan Zhang, Zhenyu Liu, Hanyu Zhang, Xinshi Chen, and Wenzhi Xiao. PXDesign: Fast, Modular, and Accurate De Novo Design of Protein Binders. *bioRxiv*, 2025. doi: 10.1101/2025.08.15.670450. URL https://www.biorxiv.org/content/10.1101/2025.08.15.670450v3. Version 3, posted 2025-12-17.

Zihan Qiu, Zekun Wang, Bo Zheng, Zeyu Huang, Kaiyue Wen, Songlin Yang, Rui Men, Le Yu, Fei Huang, Suozhi Huang, Dayiheng Liu, Jingren Zhou, and Junyang Lin. Gated Attention for Large Language Models: Non-linearity, Sparsity, and Attention-Sink-Free. In *Advances in Neural Information Processing Systems*, volume 38, pp. 100092–100118, 2025. doi: 10.52202/085713-3345. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/904e89bb4e632e75fb47f093b620b257-Abstract-Conference.html.

Danny Reidenbach, Filipp Nikitin, Olexandr Isayev, and Saee Paliwal. Applications of Modular Co-Design for De Novo 3D Molecule Generation. *arXiv preprint arXiv:2505.18392*, 2025. URL https://arxiv.org/abs/2505.18392v1.

Stephen A. Rettie, Katelyn V. Campbell, Asim K. Bera, Alex Kang, Simon Kozlov, Yensi Flores Bueso, Joshmyn De La Cruz, Maggie Ahlrichs, Suna Cheng, Stacey R. Gerben, Mila Lamb, Analisa Murray, Victor Adebomi, Guangfeng Zhou, Frank DiMaio, Sergey Ovchinnikov, and Gaurav Bhardwaj. Cyclic peptide structure prediction and design using AlphaFold2. *Nature Communications*, 16(1):4730, 2025a. doi: 10.1038/s41467-025-59940-7. URL https://doi.org/10.1038/s41467-025-59940-7.

Stephen A. Rettie, David Juergens, Victor Adebomi, Yensi Flores Bueso, Qinqin Zhao, Alexandria N. Leveille, Andi Liu, Asim K. Bera, Joana A. Wilms, Alina Üffing, Alex Kang, Evans Brackenbrough, Mila Lamb, Stacey R. Gerben, Analisa Murray, Paul M. Levine, Maika Schneider, Vibha Vasireddy, Sergey Ovchinnikov, Oliver H. Weiergräber, Dieter Willbold, Joshua A. Kritzer, Joseph D. Mougous, David Baker, Frank DiMaio, and Gaurav Bhardwaj. Accurate de novo design of high-affinity protein-binding macrocycles using deep learning. *Nature Chemical Biology*, 21(12):1948–1956, 2025b. doi: 10.1038/s41589-025-01929-w. URL https://doi.org/10.1038/s41589-025-01929-w.

Taha Rezai, Bin Yu, Glenn L. Millhauser, Matthew P. Jacobson, and R. Scott Lokey. Testing the Conformational Hypothesis of Passive Membrane Permeability Using Synthetic Cyclic Peptide Diastereomers. *Journal of the American Chemical Society*, 128(8):2510–2511, 2006. doi: 10.1021/ja0563455. URL https://doi.org/10.1021/ja0563455.

Sereina Riniker and Gregory A. Landrum. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *Journal of Chemical Information and Modeling*, 55 (12):2562–2574, 2015. doi: 10.1021/acs.jcim.5b00654. URL https://doi.org/10.1021/acs.jcim.5b00654.

Patrick J. Salveson, Adam P. Moyer, Meerit Y. Said, Gizem Gökçe, Xinting Li, Alex Kang, Hannah Nguyen, Asim K. Bera, Paul M. Levine, Gaurav Bhardwaj, and David Baker. Expansive discovery of chemically diverse structured macrocyclic oligoamides. *Science*, 384(6694):420–428, 2024. doi: 10.1126/science.adk1687. URL https://doi.org/10.1126/science.adk1687.

Dan Sindhikara, Steven A. Spronk, Tyler Day, Ken Borrelli, Daniel L. Cheney, and Shana L. Posy. Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling. *Journal of Chemical Information and Modeling*, 57(8):1881–1894, 2017. doi: 10.1021/acs.jcim.7b00052. URL https://doi.org/10.1021/acs.jcim.7b00052.

Hannes Stark, Bowen Jing, Regina Barzilay, and Tommi Jaakkola. Harmonic self-conditioned flow matching for joint multi-ligand docking and binding site design. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 46468–46494. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/stark24a.html.

Philipp Thölke and Gianni De Fabritiis. Equivariant Transformers for Neural Network based Molecular Potentials. In *International Conference on Learning Representations*, 2022. URL https://iclr.cc/virtual/2022/spotlight/6417. arXiv:2202.02541.

André 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. doi: 10.1021/acs.jcim.4c01896. URL https://doi.org/10.1021/acs.jcim.4c01896.

Elizabeth A Villar, Dmitri Beglov, Spandan Chennamadhavuni, John A Porco, Dima Kozakov, Sandor Vajda, and Adrian Whitty. How proteins bind macrocycles. *Nature Chemical Biology*, 10(9): 723–731, 2014. doi: 10.1038/nchembio.1584. URL https://doi.org/10.1038/nchembio.1584.

Vilya Research, Pascal Sturmfels, Naozumi Hiranuma, Milad Salem, Benjamin D. Sellers, Stephen Rettie, CJ San Felipe, Chase A. P. Wood, Jeffrey K. Holden, Adam P. Moyer, Patrick J. Salveson, and Ivan Anishchanka. Accurate structural modeling of chemically diverse molecular interfaces with Vilya-2. arXiv:2607.25156v1, 2026a. URL https://arxiv.org/abs/2607.25156v1.

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. *arXiv preprint arXiv:2607.09998v1*, 2026b. doi: 10.48550/arXiv.2607.09998. URL https://arxiv.org/abs/2607.09998v1. Version 1.

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. doi: 10.1021/jacs.8b13178. URL https://doi.org/10.1021/jacs.8b13178.

Nikhil Vyas, Depen Morwani, Rosie Zhao, Itai Shapira, David Brandfonbrener, Lucas Janson, and Sham Kakade. SOAP: Improving and Stabilizing Shampoo using Adam for Language Modeling. In *International Conference on Learning Representations*, pp. 93423–93444, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/e988664070e9591f93fdcf605f7dc623-Paper-Conference.pdf.

Shuzhe Wang, Jagna Witek, Gregory A. Landrum, and Sereina Riniker. Improving Conformer Generation for Small Rings and Macrocycles Based on Distance Geometry and Experimental Torsional-Angle Preferences. *Journal of Chemical Information and Modeling*, 60(4):2044–2058,

2020. doi: 10.1021/acs.jcim.0c00025. URL https://doi.org/10.1021/acs.jcim.0c00025.

Yuyang Wang, Ahmed A. 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*, volume 235 of *Proceedings of Machine Learning Research*, pp. 50400–50418, 2024. URL https://proceedings.mlr.press/v235/wang24q.html. arXiv:2311.17932.

Tina R White, Chad M Renzelman, Arthur C Rand, Taha Rezai, Cayla M McEwen, Vladimir M Gelev, Rushia A Turner, Roger G Linington, Siegfried S F Leung, Amit S Kalgutkar, Jonathan N Bauman, Yizhong Zhang, Spiros Liras, David A Price, Alan M Mathiowetz, Matthew P Jacobson, and R Scott Lokey. On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nature Chemical Biology*, 7(11):810–817, 2011. doi: 10.1038/nchembio.664. URL https://doi.org/10.1038/nchembio.664.

Jeremy Wohlwend, Gabriele Corso, Saro Passaro, Noah Getz, Mateo Reveiz, Ken Leidal, Wojtek Swiderski, Liam Atkinson, Tally Portnoi, Itamar Chinn, Jacob Silterra, Tommi Jaakkola, and Regina Barzilay. Boltz-1: Democratizing Biomolecular Interaction Modeling. *bioRxiv*, 2025. doi: 10.1101/2024.11.19.624167. URL https://www.biorxiv.org/content/10.1101/2024.11.19.624167v4. Version 4, posted 2025-05-06.

Dizhou Wu and Yike Zou. Accurate 3D Structure Prediction of Small Cyclic Peptides Containing Non-Canonical Amino Acid Residues Using an All-Atom Diffusion Model with Stereogenic Implementation. *Journal of Chemical Information and Modeling*, 66(8):4398–4408, 2026. doi: 10.1021/acs.jcim.5c03236. URL https://doi.org/10.1021/acs.jcim.5c03236.

Xuezhi Xie, Christina Z Li, Jin Sub Lee, and Philip M Kim. CyclicBoltz1, fast and accurately predicting structures of cyclic peptides and complexes containing non-canonical amino acids using AlphaFold 3 Framework. *bioRxiv*, 2025. doi: 10.1101/2025.02.11.637752. URL https://www.biorxiv.org/content/10.1101/2025.02.11.637752v1. Version 1, posted 2025-02-15.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. In *International Conference on Learning Representations*, 2022. URL https://arxiv.org/abs/2203.02923v1. arXiv:2203.02923.

Chengxuan Ying, Tianle Cai, Shengjie Luo, Shuxin Zheng, Guolin Ke, Di He, Yanming Shen, and Tie-Yan Liu. Do Transformers Really Perform Badly for Graph Representation? In *Advances in Neural Information Processing Systems*, volume 34, pp. 28877–28888, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf.

Chengyun Zhang, Wentong Wang, Renjie Zhu, Sen Cao, Yiqi Xu, Ning Zhu, Yaling Wu, Jingjing Guo, and Hongliang Duan. HighFold4: extending AlphaFold3 to accurate cyclic peptide conformation prediction via custom chemical connectivity. *Briefings in Bioinformatics*, 27(5):bbag505, 2026. doi: 10.1093/bib/bbag505. URL https://doi.org/10.1093/bib/bbag505.

Chenhao Zhang, Chengyun Zhang, Tianfeng Shang, Ning Zhu, Xinyi Wu, and Hongliang Duan. HighFold: accurately predicting structures of cyclic peptides and complexes with head-to-tail and disulfide bridge constraints. *Briefings in Bioinformatics*, 25(3):bbae215, 2024. doi: 10.1093/bib/bbae215. URL https://doi.org/10.1093/bib/bbae215.

Cheng Zhu, Sen Cao, Tianfeng Shang, Jingjing Guo, An Su, Chengxi Li, and Hongliang Duan. Predicting the structures of cyclic peptides containing unnatural amino acids by HighFold2. *Briefings in Bioinformatics*, 26(3):bbaf202, 2025. doi: 10.1093/bib/bbaf202. URL https://doi.org/10.1093/bib/bbaf202.

Jinhua Zhu, Yingce Xia, Chang Liu, Lijun Wu, Shufang Xie, Yusong Wang, Tong Wang, Tao Qin, Wengang Zhou, Houqiang Li, Haiguang Liu, and Tie-Yan Liu. Direct Molecular Conformation Generation. *Transactions on Machine Learning Research*, 2022. URL https://arxiv.org/abs/2202.01356v2. arXiv:2202.01356.

A Model and Training Specification

This appendix specifies the architecture, training objective, and inference procedure used for the reported CycPepFlow checkpoints.

A.1 Coordinates, harmonic prior, and alignment

We use a graph-conditioned harmonic prior following earlier molecular generative models (Jing et al., 2023; Stark et al., 2024; Hassan et al., 2024). For a connected molecular graph with N explicit atoms, including hydrogens, let $X \in \mathbb{R}^{N\times 3}$ store row-vector coordinates in Å and $P_N = I_N - \mathbf{1}\mathbf{1}^\mathsf{T}/N$ denote arithmetic-mean centering, without mass weighting. Let A_G be the unweighted covalent adjacency matrix and $\mathsf{L}_G = \alpha\,(diag\,(A_G\mathbf{1}) - A_G)$, with $\alpha = 1$. Given the eigendecomposition $\mathsf{L}_G = Q diag\,(\lambda_1, \ldots, \lambda_N)\, Q^\mathsf{T}$, with the translation mode listed first, we sample the prior as

$$X_0 = Q diag\left(0, \lambda_2^{-1/2}, \ldots, \lambda_N^{-1/2}\right) Z, \qquad Z_{ia} \overset{iid}{\sim} \mathcal{N}(0,1). \tag{10}$$

Here X_0 is the centered prior, I_N is the $N \times N$ identity matrix, and $\mathbf{1}$ is the length-N all-ones vector. The columns of Q are orthonormal Laplacian eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_N$; the translation eigenvalue is zero, and the remaining eigenvalues are positive. The array $Z \in \mathbb{R}^{N\times 3}$ has independent standard-normal entries Z_{ia}, with i indexing rows and a indexing Cartesian components.

Before alignment, each Cartesian column has covariance L_G^+ on the centered subspace, and the prior density on that subspace is proportional to $\exp\left[-tr\left(X_0^\mathsf{T}\mathsf{L}_G X_0\right)/2\right]$. This prior depends only on covalent connectivity. The dataset loader excludes disconnected molecular records. The superscript in L_G^+ denotes the Moore–Penrose pseudoinverse, and tr denotes the matrix trace.

During training, we align the sampled prior to the observed conformer using a proper rotation. For centered endpoints, the alignment is

$$R^\star = \underset{R\in SO(3)}{argmin} \parallel X_0 R^\mathsf{T} - X_1 \parallel_F^2, \qquad X_0 \leftarrow X_0\,(R^\star)^\mathsf{T}. \tag{11}$$

Here X_1 is the observed conformer, $R \in SO(3)$ is a proper three-dimensional rotation, $R^\star$ is the minimizing rotation, and $\|\cdot\|_F$ is the Frobenius norm.

The Kabsch alignment includes a determinant correction to exclude reflections. Both endpoints are centered before interpolation, so Equation (11) specifies the alignment used by the training path without a translational offset.

A.2 Noisy conditional path and velocity-regression objective

For each molecule, we independently sample a flow time $t \sim \mathcal{U}\left[10^{-4}, 0.9999\right]$ and centered Gaussian noise $\epsilon = P_N Z'$. With base noise amplitude $\sigma = 0.1$, the noisy conditional path and its samplewise velocity are

$$\sigma_t = \sigma\sqrt{t(1-t)}, \qquad \dot{\sigma}_t = \frac{\sigma(1-2t)}{2\sqrt{t(1-t)}}, \tag{12}$$

$$X_t = (1-t)X_0 + tX_1 + \sigma_t\epsilon, \qquad u_t = X_1 - X_0 + \dot{\sigma}_t\epsilon. \tag{13}$$

Here t is dimensionless, $\mathcal{U}$ denotes the uniform distribution, and $Z^{'}$ has independent standard-normal entries before centering. For fixed endpoints and noise, u_t is the time derivative of X_t, and $\dot{\sigma}_t$ is the derivative of the noise amplitude σ_t. The restricted training-time interval avoids the endpoint singularities in $\dot{\sigma}_t$. In this subsection, X_0 denotes the aligned prior and X_1 the centered observed conformer.

The velocity loss in Equation (3) averages the unsquared Euclidean error over atoms within each molecule and then over molecules in the local minibatch:

$$\mathcal{L}_{vel} = \frac{1}{B}\sum_{g=1}^{B}\frac{1}{N_g}\sum_{i=1}^{N_g} \parallel v_{\theta,gi}\left(X_{t_g}, G_g, t_g\right) - u_{t_g,i} \parallel_2 . \tag{14}$$

Here B is the local minibatch size, g indexes molecules, N_g is the atom count of molecule g, and i indexes its atoms. Because the norm in Equation (14) is not squared, this objective differs from standard squared-loss conditional flow matching (Lipman et al., 2023). The standard squared-loss conditional-to-marginal gradient-equivalence result is not invoked for this unsquared objective.

A.3 Conditioning and graph construction

Atomic number is encoded by a learned embedding table. Each atom also has a two-component descriptor containing its RDKit chirality-category index and formal-charge-category index. Chirality categories are unspecified, tetrahedral CW, tetrahedral CCW, other, and miscellaneous; formal-charge categories are $-5, -4, \ldots, 5$, followed by miscellaneous. The category indices are supplied as floating-point inputs, rather than one-hot vectors or physical charge values, to

$$a_i = W_2\mathrm{LN}\left(SiLU\left(W_1 f_i + b_1\right)\right) + b_2, \qquad a_i \in \mathbb{R}^d. \tag{15}$$

Here f_i is the two-component descriptor of atom i, a_i is its attribute embedding, and d is the feature width. The parameters W_1, W_2 and b_1, b_2 are learned weights and biases. LN denotes LayerNorm applied independently to each atom across feature channels, and $SiLU$ denotes the sigmoid linear unit. The encoder widths are $2 \to d \to d$, with SiLU then LayerNorm between the two linear maps.

RDKit's tetrahedral CW/CCW tag encodes orientation relative to the stored neighbor order, not an absolute R/S or L/D label.

Training records preserve explicit-H coordinates and atom-mapped isomeric SMILES. The loader retains hydrogens, restores the mapped atom order, and derives node descriptors, covalent edges, and chirality tensors from the reconstructed molecule. The input-SMILES helper instead adds hydrogens before featurization. Batching offsets center and neighbor indices together. The auxiliary triple-product target is recomputed from the reference coordinates using the same index-sorted neighbor triple as the prediction (Appendix A.7). Raw CW/CCW categories are not invariant under arbitrary atom reindexing or re-serialization; features and index-dependent tensors must be regenerated consistently.

At each vector-field evaluation, the directed interaction graph combines covalent, exact two-hop, exact three-hop, and current-coordinate radius pairs, with relation codes $1, 2, 3, 0$, respectively. Topological classes exclude self-pairs and pairs assigned to a shorter hop. Radius neighbors are queried within 10 Å using torch-cluster radius_graph with max_num_neighbors=32 and no self-loops before taking the union; overlapping pairs retain their topological code. The final interaction graph may therefore contain more than 32 incoming neighbors per atom. Runtime relation codes do not encode chemical bond order. All 20 layers share the interaction graph and geometric edge features at the current coordinate state. After each ODE update, the next vector-field evaluation recomputes radius neighbors and geometric features while retaining the covalent connectivity and hop definitions.

Geometric edge features. For an incoming edge $j \to i$, the implementation uses

$$\Delta_{ij} = X_{t,j} - X_{t,i}, \qquad \rho_{ij} = \parallel \Delta_{ij} \parallel_2^2, \qquad \delta_{ij} = \frac{\Delta_{ij}}{\max\left(\rho_{ij}, 10^{-2}\right)}. \tag{16}$$

For an incoming edge $j \to i$, j is the sender and i the receiver, with coordinates $X_{t,j}, X_{t,i}$. The displacement Δ_{ij} has squared length ρ_{ij}. Its scaled form δ_{ij} uses a denominator floor of 10^{-2} Å² and is not a unit vector.

The radial basis consists of 64 trainable exponential-normal features evaluated at the squared distance ρ_{ij}:

$$C(\rho) = \frac{1}{2}\left[1 + cos(\pi\rho/c)\right] \mathbf{1}\{\rho < c\}, \qquad c = 10, \tag{17}$$

$$\phi_k(\rho) = C(\rho) \exp\left[-\beta_k \left(\exp(-5\rho/c) - \mu_k\right)^2\right], \tag{18}$$

$$f_{ij} = \left[\phi_1\left(\rho_{ij}\right), \ldots, \phi_{64}\left(\rho_{ij}\right), \tau_{ij}\right], \qquad \tau_{ij} \in \{0, 1, 2, 3\}. \tag{19}$$

Here $C(\rho)$ is the cosine envelope evaluated at squared separation ρ; the numerical cutoff $c = 10$ is in squared-coordinate units (Å²). Each radial function ϕ_k, for $k = 1, \ldots, 64$, has a trainable center μ_k and width coefficient β_k. The edge descriptor f_{ij} concatenates these radial features with the scalar relation code τ_{ij}: 0 for radius-only pairs and 1, 2, or 3 for the exact covalent-hop classes. The centers are initialized as $linspace\left(e^{-10}, 1, 64\right)$, and each width coefficient as $\left[(2/64)\left(1 - e^{-10}\right)\right]^{-2}$. Both are directly trainable; the exponential transform in Equation (18) uses $5/c = 1/2$.

Consequently, the radial and scalar-attention envelope vanishes at a physical separation of $\sqrt{10} \approx 3.16$ Å, although the radius query extends to 10 Å. Vector messages are not multiplied by this envelope and may remain nonzero beyond its cutoff through the relation-code input and biased edge projections.

For atomic numbers Z_i, neighbor initialization uses independent $100 \times d$ embedding tables E_0 and E_n. Excluding self-pairs, it forms $m_i = \sum_{j \to i} E_n\left(Z_j\right) \odot \{C\left(\rho_{ij}\right)\left(W_r f_{ij} + b_r\right)\}$, then $h_i^{(0)} = W_c\left[E_0\left(Z_i\right), m_i\right] + b_c$. The affine maps have widths $65 \to d$ and $2d \to d$, respectively, with no following activation. The outer envelope multiplies the entire projected edge descriptor, in addition to the envelope already inside its radial entries. This is concatenation followed by projection, not an extra residual addition.

A.4 Local scalar–vector block

The local block extends the TorchMD scalar–vector backbone (Thölke and De Fabritiis, 2022). Each atom carries a scalar stream $h_i \in \mathbb{R}^d$, invariant under proper rotations, and a corresponding equivariant vector stream $V_i \in \mathbb{R}^{3\times d}$. Atomic-number and neighbor embeddings initialize the scalar stream; vector channels are initialized to zero. Let (h, V) denote the inputs selected by the attention-residual mechanism in Appendix A.6. The block first computes

$$z_i = \text{LN}\left(MLP\left([h_i, t, a_i]\right)\right). \tag{20}$$

A multilayer perceptron (MLP), followed by LayerNorm, transforms the concatenated scalar state h_i, raw time t, and attributes a_i into the locally conditioned feature $z_i \in \mathbb{R}^d$. The MLP has widths $2d + 1 \to d \to d$, with SiLU between its linear maps and LayerNorm afterward; no separate time embedding is used.

Query and key projections are followed by LayerNorm and split into H heads of width $d_h = d/H$. Linear–SiLU maps of the edge descriptor produce channelwise key and value factors D_{ij}^K and D_{ij}^U. For head m,

$$A_{ij}^{(m)} = SiLU\left(\sum_c q_{ic}^{(m)} k_{jc}^{(m)} D_{ijc}^{K,(m)}\right) C\left(\rho_{ij}\right), \tag{21}$$

$$\left(b_{ij}^{(0,m)}, b_{ij}^{(1,m)}, b_{ij}^{(2,m)}, b_{ij}^{(3,m)}\right) = split\left(Linear_U\left(z_j\right)^{(m)} \odot D_{ij}^{U,(m)}\right), \tag{22}$$

Here m indexes attention heads and c indexes channels within a head. The normalized query and key components are $q_{ic}^{(m)}$ and $k_{jc}^{(m)}$, and $D_{ij}^{K,(m)}$ and $D_{ij}^{U,(m)}$ are edge-derived modulation factors. The scalar weight $A_{ij}^{(m)}$ modulates the scalar message. The learned value projection $Linear_U$ and channel partition $split$ produce four d_h-component coefficient arrays $b_{ij}^{(0,m)}, \ldots, b_{ij}^{(3,m)}$. Before head splitting, the edge key and value maps have widths $65 \to d$ and $65 \to 4d$, respectively. Each is a

single affine map followed by SiLU, with no hidden layer. The $d \to d$ query and key projections are normalized across all d channels before head reshaping.

$$m_{ij}^{h,(m)} = A_{ij}^{(m)} b_{ij}^{(0,m)}, \tag{23}$$

$$m_{ij}^{V,(m)} = V_j^{(m)} \odot b_{ij}^{(1,m)} + \delta_{ij} \otimes b_{ij}^{(2,m)} + \left(\delta_{ij} \times V_j^{(m)}\right) \odot b_{ij}^{(3,m)}. \tag{24}$$

Here $m_{ij}^{h,(m)}$ and $m_{ij}^{V,(m)}$ are scalar and vector messages, and $V_j^{(m)}$ is the sending atom's $3 \times d_h$ vector-head slice. The operators $\odot$, $\otimes$, and $\times$ denote channelwise multiplication with spatial broadcasting, a spatial-by-channel outer product, and a channelwise cross product, respectively.

Local attention uses neither a neighbor softmax nor a $1/\sqrt{d_h}$ scaling factor. The vector message in Equation (24) is not multiplied by the scalar attention weight $A_{ij}^{(m)}$. The cross-product construction is equivariant under proper rotations, but reflection equivariance is not imposed.

Messages are summed over incoming edges. Let $U_i^{(1)}, U_i^{(2)}, U_i^{(3)}$ be three learned, bias-free channel projections of V_i, and let $\left(o_i^{(1)}, o_i^{(2)}, o_i^{(3)}\right)$ denote the three outputs of a linear projection of the concatenated scalar-message sums. The local residual increments are

$$c_{ic}^{\text{vec}} = \sum_{a=1}^{3} \left(U_i^{(1)}\right)_{ac} \left(U_i^{(2)}\right)_{ac}, \tag{25}$$

$$\Delta h_i = c_i^{\text{vec}} \odot o_i^{(2)} + o_i^{(3)}, \tag{26}$$

$$\Delta V_i = \left(U_i^{(3)}\right) \odot o_i^{(1)} + \sum_j m_{ij}^{V}. \tag{27}$$

Here a and c index Cartesian components and feature channels, respectively, and c_i^{vec} collects the channelwise contractions c_{ic}^{vec}. The projected vectors $U_i^{(1:3)}$ have shape $3 \times d$, and each scalar coefficient array in $o_i^{(1:3)}$ has d channels. In Equation (27), the sum over j aggregates incoming vector messages after head concatenation.

The local outputs are $h_i + \Delta h_i$ and $V_i + \Delta V_i$.

A.5 Gated scalar global self-attention and shortest-path bias

Every local block is followed by molecule-wide scalar self-attention with head-specific output gating adapted from Qiu et al. (2025). Normalize the post-local scalar state as z_i=LN (h_i), and define $q_i^{(m)} = W_Q^{(m)} z_i$, $k_i^{(m)} = W_K^{(m)} z_i$, and $u_i^{(m)} = W_U^{(m)} z_i$. Head m computes

$$\ell_{ij}^{(m)} = \frac{\left(q_i^{(m)}\right)^{\mathsf{T}} k_j^{(m)}}{\sqrt{d_h}} + B_{ij}^{(m)}, \qquad B_{ij}^{(m)} = b_{\ell,m}\left[\kappa(i,j)\right], \tag{28}$$

The learned maps $W_Q^{(m)}, W_K^{(m)}, W_U^{(m)}$ produce d_h-channel queries $q_i^{(m)}$, keys $k_i^{(m)}$, and values $u_i^{(m)}$. The logit $\ell_{ij}^{(m)}$ connects query atom i to key atom j in head m. Its additive bias $B_{ij}^{(m)}$ uses the layer-specific lookup table $b_{\ell,m}$ and path bin $\kappa(i,j)$.

$$A_{ij}^{(m)} = \frac{\exp \ell_{ij}^{(m)}}{\sum_{k \in \mathcal{V}_g} \exp \ell_{ik}^{(m)}}, \qquad y_i^{(m)} = \sum_{j \in \mathcal{V}_g} A_{ij}^{(m)} u_j^{(m)}, \tag{29}$$

$$g_i^{(m)} = 2sigmoid\left(W_g z_i + b_g\right)_m, \qquad \widetilde{y}_i^{(m)} = g_i^{(m)} y_i^{(m)}. \tag{30}$$

Here $\mathcal{V}_g$ is the atom set of molecule g, and j, k index its key/value atoms. The softmax weights $A_{ij}^{(m)}$ produce attended values $y_i^{(m)}$, which are multiplied by the rescaled sigmoid gates $g_i^{(m)}$ to obtain $\widetilde{y}_i^{(m)}$. The learned parameters W_g and b_g produce one gate per atom and head, with values in (0, 2).

Attention includes all atom pairs within a molecule, including self-pairs and local neighbors. Padded keys are masked, and padded query outputs are zeroed. Concatenated gated heads pass through an output linear map and a two-linear-layer SiLU MLP before residual addition to the scalar stream. The final MLP projection is zero-initialized. The gate projection is also zero-initialized, giving an initial multiplier of one because of the factor of two in Equation (30). This branch leaves the vector stream unchanged.

The optional all-pairs graph-geodesic bias (APG) uses categorical covalent shortest-path lengths. Let $d_G(i, j)$ be the shortest-path length in the symmetrized type-1 covalent graph. The corresponding bin is

$$\kappa(i,j) = \begin{cases} \min\left(d_G(i,j), 32\right), & i, j \text{ connected}, \\ 33, & i, j \text{ disconnected or padded}. \end{cases} \tag{31}$$

The bin $\kappa(i, j)$ is zero for self-pairs, 32 for connected pairs with shortest-path length at least 32, and 33 for disconnected or padded pairs. When APG is disabled, all pair biases are set to zero ($B = 0$).

Each of the 20 layers has a separate $34 \times H$ bias table initialized to zero, adding 5,440 parameters for B and 8,160 for L. APG supplements the topology already encoded by local message passing and does not use coordinate distances. The implemented dense Floyd–Warshall calculation has $O\left(N^3\right)$ time complexity and is repeated once per vector-field evaluation; its result is reused across all layers. Global attention itself is quadratic in the atom count N.

A.6 Full-depth AttnRes and Conditional ProductLite

Full-depth AttnRes mixes the initial post-neighbor-embedding state and all completed block outputs. A block output is stored only after global attention and any scheduled Conditional ProductLite update. At block ℓ, each available source $s < \ell$ receives an atomwise weight

$$\alpha_{\ell,s,i} = softmax_s\left(\frac{q_\ell^{\mathsf{T}} \mathrm{LN}\left(h_{s,i}\right)}{\sqrt{d}}\right). \tag{32}$$

The same weights mix both streams, giving $\widetilde{h}_{\ell,i} = \sum_s \alpha_{\ell,s,i} h_{s,i}$ and $\widetilde{V}_{\ell,i} = \sum_s \alpha_{\ell,s,i} V_{s,i}$. These mixtures are both the inputs and the residual baselines for the next local block. Twenty block-specific queries and one final-readout query are learned. Query entries are initialized from a zero-mean Gaussian with standard deviation $d^{-1/2}$, yielding approximately uniform initial mixing. The depth softmax is evaluated in float32. The final readout mixes the initial state and all 20 completed block outputs, then applies LayerNorm to the scalar stream.

Here ℓ is the destination block, s is a source depth, i indexes atoms, and $q_\ell \in \mathbb{R}^d$ is the learned query shared across atoms. The weights $\alpha_{\ell,s,i}$ are normalized over source depth separately for each atom; the same weights are applied to the scalar and vector states $h_{s,i}, V_{s,i}$.

Conditional ProductLite is a lightweight MACE-inspired residual applied after global attention in blocks 5, 10, 15, and 20. With channelwise vector powers $r_{ic} = \sum_{a=1}^{3} V_{iac}^2$ and scalar–vector-power products $p_i = h_i \odot r_i$, its conditioning vector is

$$\xi_i = \left[\mathrm{LN}\left(h_i\right), \mathrm{LN}\left(r_i\right), \mathrm{LN}\left(p_i\right), \mathrm{LN}\left(a_i\right), t\right], \tag{33}$$

Here r_{ic} is the squared spatial norm of vector channel c, with a indexing Cartesian components. The arrays r_i and p_i contain vector powers and their channelwise products with the scalar state h_i,

respectively. The conditioning vector ξ_i concatenates four separately normalized d-channel arrays and raw time t, for a total of $4d + 1$ components.

$$\left(u_i, g_i^h, g_i^V\right) = W_o SiLU\left(W_i \xi_i + b_i\right) + b_o, \tag{34}$$

$$\Delta h_i^{\text{prod}} = u_i \odot sigmoid\left(g_i^h\right), \qquad \Delta V_i^{\text{prod}} = V_i \odot tanh\left(g_i^V\right). \tag{35}$$

Here W_i, b_i and W_o, b_o denote the input and output projection parameters, respectively; the subscripts indicate projection roles rather than atom indices. The three d-channel outputs are the scalar proposal u_i and gate inputs g_i^h, g_i^V, which define the residuals $\Delta h_i^{\text{prod}}, \Delta V_i^{\text{prod}}$.

The final projection W_o, b_o is zero-initialized, so both ProductLite residuals initially vanish. The invariant product $h \odot \parallel V \parallel^2$ uses vector norms computed over Cartesian components separately for each channel. This is a lightweight product construction, not a full Clebsch–Gordan tensor-product expansion or a MACE potential.

After the final AttnRes mixture and scalar normalization, a two-block vector readout produces one three-vector per atom. Subtracting the mean predicted vector over atoms gives the centered velocity field $v_\theta \in \mathbb{R}^{N\times 3}$, parameterized by the learned network and readout parameters θ.

The two gated readout blocks reduce vector channels $d \to d/2 \to 1$. Each concatenates channelwise norms of one bias-free vector projection with the scalar stream; its two-linear-layer SiLU MLP supplies scalars and gates for a second bias-free vector projection. The MLP widths are $2d \to d \to d$ and $d \to d/2 \to d + 1$. Block 1 splits its outputs into $d/2$ scalars and gates, with an additional SiLU on the scalar half; block 2 splits into d discarded scalars and one gate, with no final scalar activation. Neither readout block uses LayerNorm. The norm-safety mask evaluates norm features only when all components of the first projected vector are nonzero; otherwise those features are zero. This componentwise zero guard depends on Cartesian axes and can break exact rotation equivariance at mask-changing degeneracies. This implementation qualification does not establish any effect on the reported benchmark results. The resulting vector is centered as specified above.

A.7 Chirality supervision and masking

Inspired by Vilya-1 (Vilya Research et al., 2026b), we use an auxiliary stereochemical objective based on signed scalar triple products. Eligible centers are source-specified RDKit tetrahedral CW/CCW atoms with four explicit neighbors. We sort the four neighbor indices and select the first three, denoted j_1, j_2, j_3, using the same ordered triple for reference and predicted coordinates. Define

$$\nu_i(X) = \left(\left(X_{j_1} - X_i\right) \times \left(X_{j_2} - X_i\right)\right) \cdot \left(X_{j_3} - X_i\right), \tag{36}$$

$$\bar{\nu}_i(X) = \frac{\nu_i(X)}{\max\left(\prod_{m=1}^{3} \parallel X_{j_m} - X_i \parallel_2, \varepsilon\right)}, \qquad \varepsilon = 10^{-7}. \tag{37}$$

For center i, X_i and X_{j_m} are the center and selected-neighbor coordinates, with $m = 1, 2, 3$. The raw scalar triple product $\nu_i(X)$ measures signed orientation, and $\bar{\nu}_i(X)$ is its length-normalized form. The implementation uses the same numerical tolerance $\varepsilon = 10^{-7}$ as a length-product floor in Å³ in Equation (37), and a dimensionless reference-volume threshold in Equation (38). It also floors the dimensionless scale denominator in Equation (39); with $\eta = 1$, that denominator is exactly one.

The target sign is obtained from the geometry of the currently sampled observed conformer, not directly from the stored CW/CCW sign tensor. Let

$$\mathcal{C} = \{i : |stopgrad\left(\bar{\nu}_i\left(X_1\right)\right)| > \varepsilon\}, \qquad s_i = sign\left(stopgrad\left(\bar{\nu}_i\left(X_1\right)\right)\right). \tag{38}$$

Here $\mathcal{C}$ contains eligible centers whose normalized reference-volume magnitude exceeds ε, and s_i is the corresponding orientation sign. The $stopgrad$ operator prevents differentiation through the reference.

Using the local endpoint extrapolation $\widehat{X}_1 = X_t + (1-t)v_\theta\left(X_t, G, t\right)$, define the auxiliary loss as

$$\mathcal{L}_{chiral} = \begin{cases} \frac{1}{|\mathcal{C}|}\sum_{i\in\mathcal{C}} softplus\left(-\frac{s_i\bar{\nu}_i\left(\widehat{X}_1\right)}{\max(\eta,\varepsilon)}\right), & |\mathcal{C}|>0, \\ 0, & |\mathcal{C}|=0, \end{cases} \qquad \eta = 1. \tag{39}$$

Here $\widehat{X}_1$ is the local endpoint estimate, $\eta = 1$ sets the loss scale, and $softplus$ is the softplus activation. The loss $\mathcal{L}_{chiral}$ averages over the $|\mathcal{C}|$ valid center instances in a local minibatch, with i including molecule membership, rather than first averaging within each molecule. An empty valid-center set or missing center tensors gives zero auxiliary loss.

The auxiliary loss is applied at all sampled times, without an additional positive margin, predicted-volume mask, or explicit finite-value mask. It remains positive even when the predicted sign is correct. The endpoint estimate is a local extrapolation, not an exact denoising formula for the noisy path. Training uses $\mathcal{L} = \mathcal{L}_{vel} + 0.5\mathcal{L}_{chiral}$ from the first step, with no chirality warmup or weight ramp.

Here $\mathcal{L}$ is the total loss and the chirality weight is fixed at 0.5. Distributed data parallelism averages rank-local gradients; this is not generally equivalent to weighting all valid centers globally by their counts when ranks contain different numbers of valid centers.

A.8 Sizes, optimization, and inference

Table 1: Model sizes and training layouts. "Devices × batch" gives the device count and per-device batch size. Selected epochs are the zero-based indices stored in the checkpoints. All variants have 20 layers and an effective global batch size of 64.

Variant	Width	Heads	Parameters	Devices × batch	Epoch selected
B	192	8	18,303,745	4×16	144
APG-B	192	8	18,309,185	4×16	144
L	300	12	43,422,219	4×16	144
APG-L	300	12	43,430,379	8×8	185

All four variants were trained from scratch for 200 epochs with seed 3221. SOAP uses a learning rate of 8×10^{-4}, moment-decay coefficients $(0.95, 0.95)$, weight decay 10^{-8}, and stabilization constant $\epsilon = 10^{-8}$. The preconditioner is updated every 10 steps, with a maximum preconditioned dimension of 4096, one-dimensional preconditioning disabled, and bias correction enabled. The cosine warmup/restart schedule uses 8,000 warmup optimizer-step calls, an 80,000-call first cycle including warmup, a minimum learning rate of 10^{-5}, a cycle multiplier of one, and a cycle decay factor of 0.2. The schedule advances once per optimizer-step call, without gradient accumulation. SOAP's first call initializes preconditioners without changing parameters but still advances the schedule.

Before each optimizer step, the whole-model gradient L2 norm is clipped at $T = min\{1.5\mu_H + 2\sigma_H, 100\}$, where μ_H and σ_H are the history mean and population standard deviation. The history holds at most 50 entries, initialized with the single value 3000. For pre-clipping norm g, the inserted entry is T if $T < g < 100$, and g otherwise; the oldest entry is removed when needed. Thus gradients with norms at least 100 are clipped, but their unmodified norms enter the history. This is whole-model norm clipping, not parameter-wise adaptive gradient clipping.

For both training and validation, the loader retains up to 30 conformers per molecule and samples one retained conformer at each molecule access. It does not enumerate all retained conformers in every epoch.

Validation monitors the total objective $\mathcal{L}_{vel} + 0.5\mathcal{L}_{chiral}$, not velocity loss alone; the prior, time, and path noise are also resampled. Epoch logging weights each rank-local minibatch loss by its molecule count and synchronizes numerator and denominator across ranks. The chirality component

is first averaged over valid centers within that local minibatch, so its epoch value is not a globally center-weighted mean. Every epoch, the five smallest total-validation-loss checkpoints are retained, alongside separate rolling recovery checkpoints saved every 500 training steps and every epoch. The selected checkpoints in Table 1 agree with the minima over all 200 logged validation epochs; selection does not use test-set COV, MAT, or stereochemical accuracy.

At inference, we draw a fresh centered harmonic prior and integrate the learned velocity field using 50 explicit Euler steps:

$$X^{k+1} = X^k + \frac{1}{50} v_\theta \left(X^k, G, k/50 \right), \qquad k = 0, \ldots, 49. \tag{40}$$

Here X^k is the coordinate state at step k, initialized from the sampled harmonic prior; G is the fixed molecular graph, and θ denotes the trained parameters. The Euler updates are deterministic conditional on the sampled initial coordinates.

The ODE state is stored in float32. CUDA network evaluations use BF16 autocast, except that the scalar global-attention branch is explicitly evaluated in float32 by the generation launcher. This inference procedure applies no target alignment, stochastic churn, post-hoc global reflection, or force-field relaxation. Generation batch sizes are 192 for B/APG-B and 96 for L/APG-L. During training, APG-L uses eight data-parallel processes rather than four; the shared global batch size therefore does not imply identical data-parallel layouts.

The released YAML files are inference configurations rather than unchanged copies of the original training files: they omit logging and checkpoint callbacks, replace internal paths, and explicitly specify selected constructor defaults. The settings reported here were cross-checked against the original run configurations, source-checksum records, completed-training logs, and checkpoint tensors. The radius-query cutoff for the reported runs is 10 Å, as specified in Appendix A.3.

B Block-Level Model Diagrams

Shared notation and implementation details. Figures 5–8 use the notation defined in Appendix A: scalar features $h_i \in \mathbb{R}^d$, vector features $V_i \in \mathbb{R}^{3\times d}$, atom attributes a_i, and raw time t. The channel width is d, the number of attention heads is H, and the per-head width is $d_h = d/H$. The B and APG-B variants use $d = 192, H = 8$, whereas the L and APG-L variants use $d = 300, H = 12$. All four variants have 20 blocks. In Figure 5, s indexes source depth.

The coordinate symbols x_t and r_i in the diagrams denote X_t and its atom row $X_{t,i}$, respectively. By contrast, r_{ic} in Figure 8 denotes the squared vector magnitude for each atom and channel, not a coordinate. The symbol d_{ij} denotes interatomic distance, and τ is shorthand for τ_{ij}. The symbol z_i denotes locally conditioned scalar features in Figure 6 and normalized scalar features after the local update in Figure 7. GGA denotes gated global attention.

The contraction $\langle \cdot, \cdot \rangle_{xyz}$ sums over Cartesian components separately for each channel. The operators $\odot, \otimes, \times$ follow the channelwise conventions specified in Appendix A.4. Vector-channel projections are bias-free, and normalization is not applied directly to Cartesian vector components.

In Figure 7, $Embed_m\left(d_G\right)$ denotes the bias lookup for layer ℓ and head m, written $b_{\ell,m}\left[\kappa(i,j)\right]$, using binned covalent shortest-path distances. The symbol u denotes scalar value features.

In Figure 8, $h \odot \sum_{xyz} V^2$ denotes the channelwise product of scalar features and squared vector magnitudes. After residual addition, the total multiplier for each vector channel is $1 + tanh\left(g^V\right)$, where g^V is the corresponding scalar gate input. In Figures 7 and 8, σ denotes the logistic sigmoid function.

C Component and Optimizer Ablations

Figure 9 compares CycPepFlow-B with seven variants, including the joint five-strip (AdamW) configuration. All ablation experiments use the same training controls as the parent model, except for the explicitly stated component or optimizer changes, enabling direct assessment of their effects.

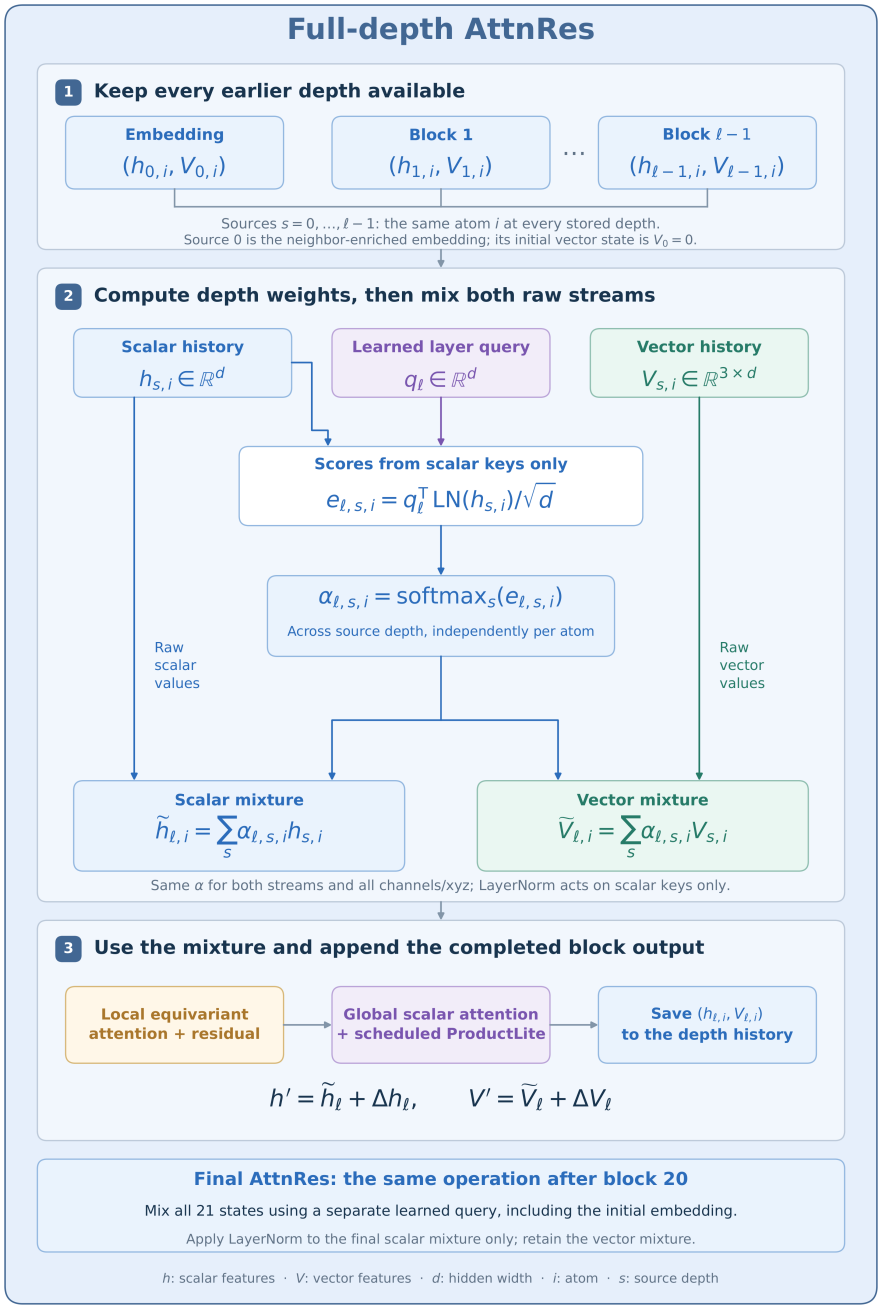


Figure 5: **Full-depth AttnRes.** For each atom, depth weights computed from scalar keys mix the raw scalar and vector histories. The same weights are used for both streams, and the resulting mixtures serve as the inputs and residual baselines for the local update. The completed block output is appended to the depth history only after all scheduled updates. A final AttnRes operation mixes all 21 stored states, including the initial embedding, after block 20. Invariant scalar weights preserve rotational equivariance.

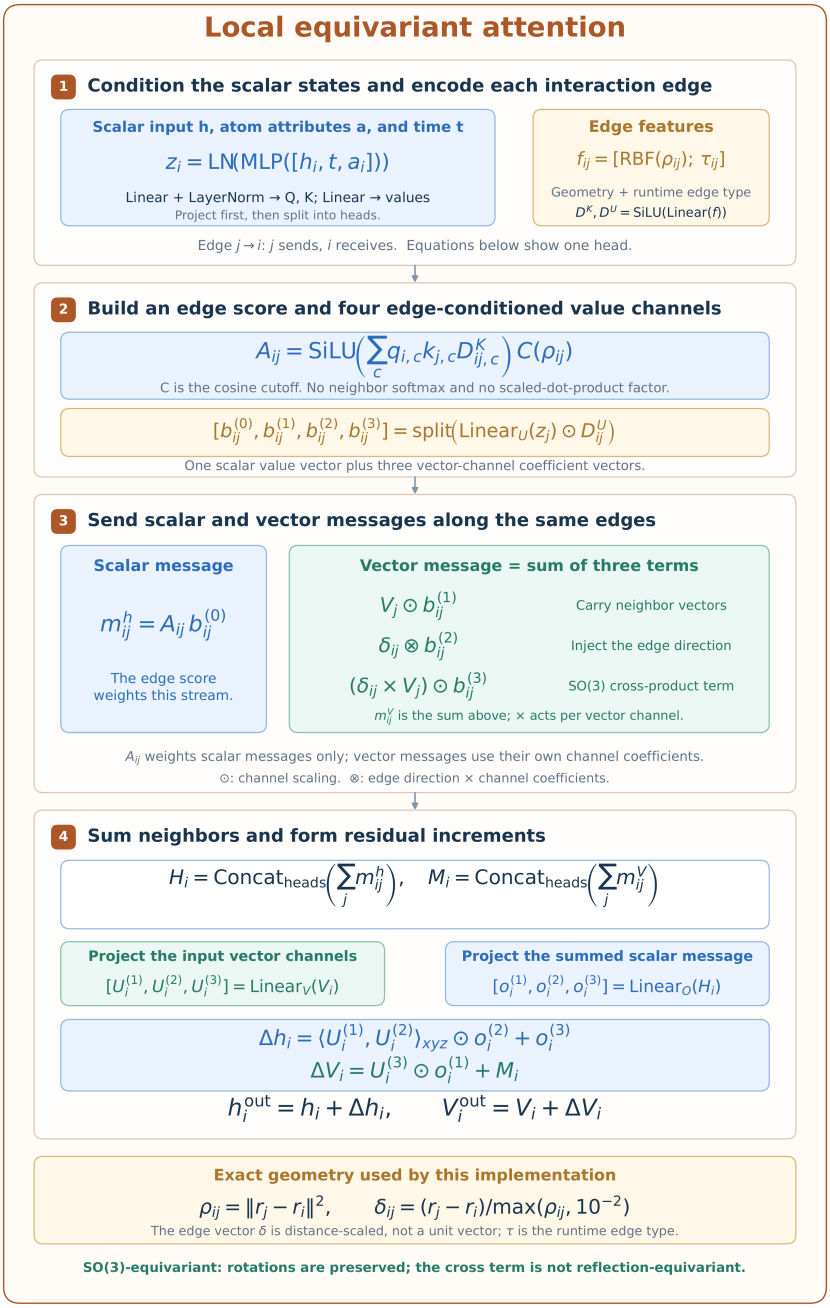


Figure 6: **Local equivariant attention.** Follow stages 1–4 from top to bottom. This is the local update within Figure 5, not an additional residual. The squared-distance geometry and vector-message operators are retained exactly; U^{(1)}, U^{(2)}, U^{(3)} denote the three projected vector-channel arrays.

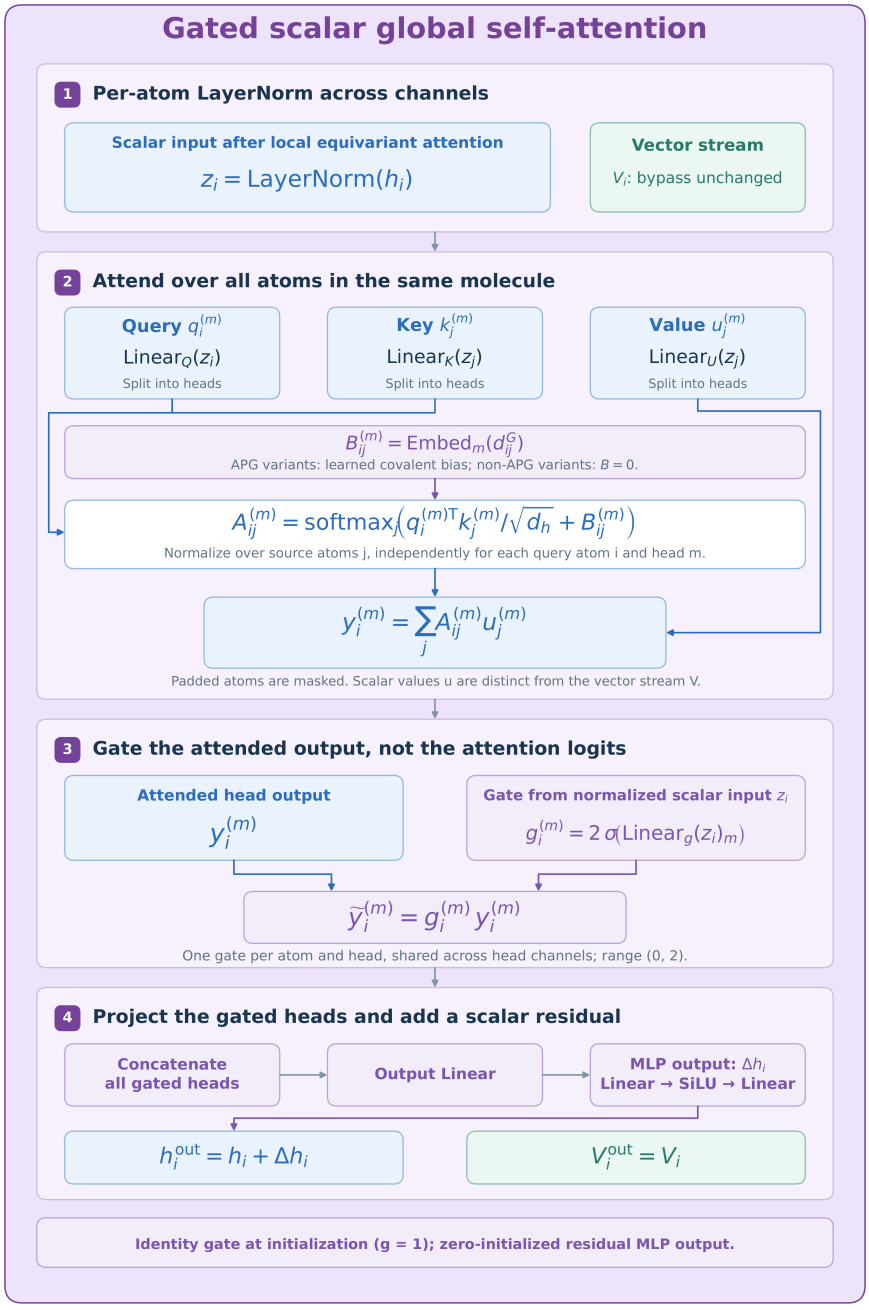


Figure 7: **Gated scalar global self-attention.** LayerNorm acts across channels for each atom, not across atoms. Attention includes self and local pairs within the same molecule. APG uses layer-specific binned covalent shortest-path bias; gates act on attended values, not logits. This branch follows every local layer, leaving the vector stream unchanged.

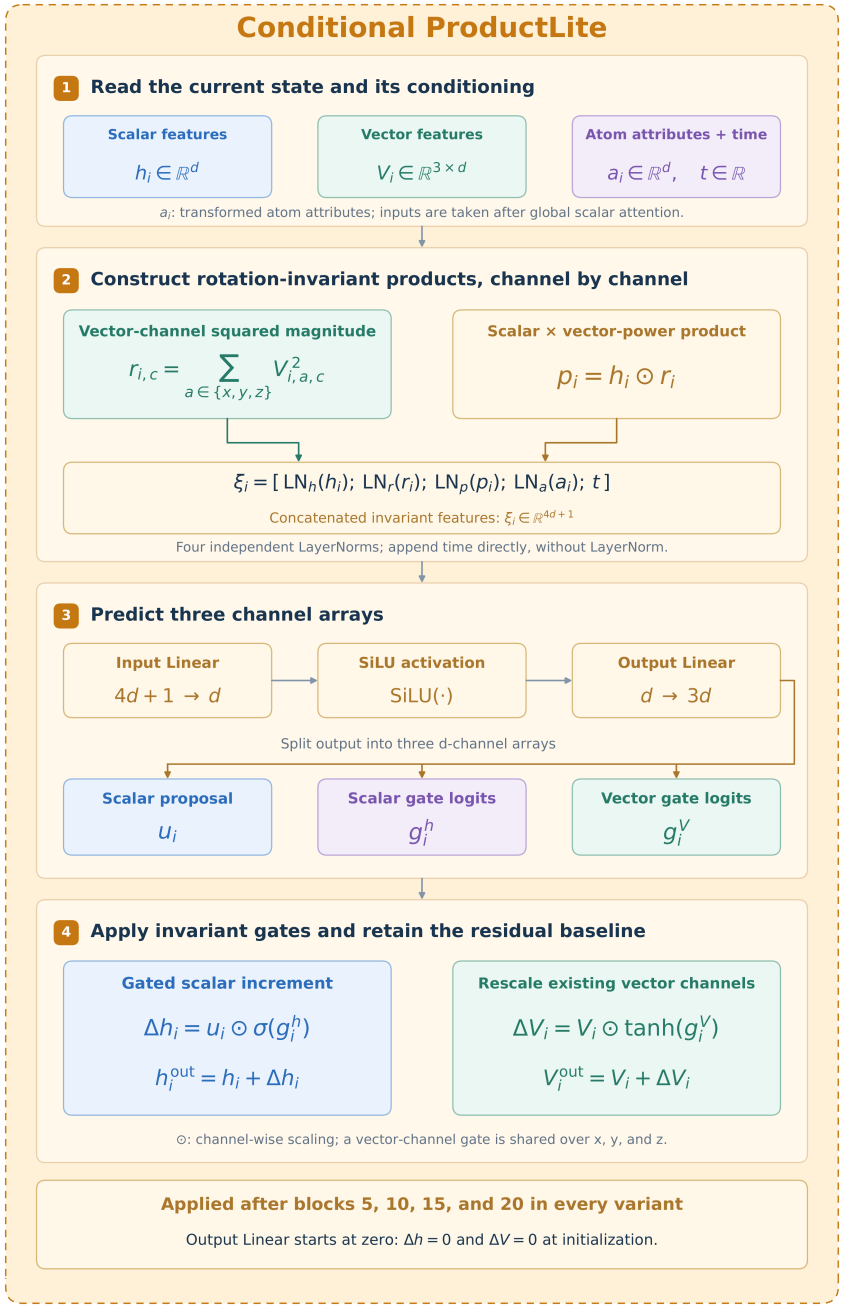


Figure 8: **Conditional ProductLite.** Read the invariant-product and gated-residual construction from top to bottom. The three d-channel outputs are scalar arrays, not three spatial vectors. The dashed border denotes execution at blocks 5, 10, 15, and 20 in each of the four main variants. These lightweight invariant products rescale existing vectors rather than forming a full MACE tensor-product expansion.

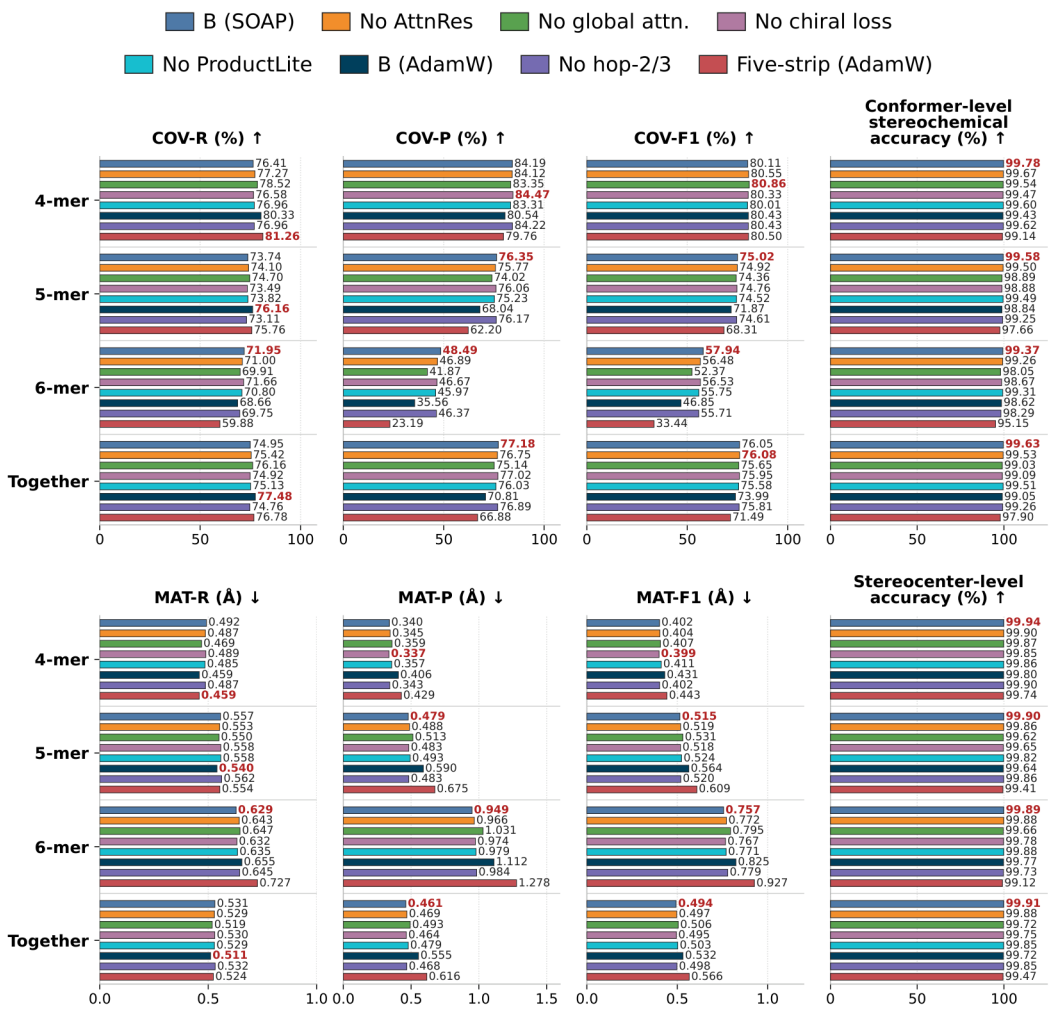


Figure 9: Component and optimizer ablations of CycPepFlow-B. "B (SOAP)" is the full parent model; "B (AdamW)" replaces SOAP with AdamW at the same learning rate. "No chiral loss" retains chirality conditioning while removing the stereochemical loss. "No hop-2/3" jointly removes the explicit two- and three-hop interaction classes while retaining covalent and radius edges; former hop-2/3 pairs can therefore remain as radius-only pairs. "Five-strip (AdamW)" removes AttnRes, gated global attention, ProductLite, chirality loss, and explicit hop-2/3 interactions while replacing SOAP with AdamW. All variants use the same training controls as the parent model except for the explicitly stated architectural, loss, or optimizer changes. "Together" uses metric-specific pooling; red/bold numerical labels mark split-best values. Tables 7–10 contain the complete numerical results.

Appendix F provides the complete split-wise results, and Table 11 reports aggregate differences relative to the parent model.

Global attention and product residuals primarily affect precision. Removing gated scalar global self-attention increases COV-R by 1.21 percentage points (pp), but decreases COV-P by 2.05 pp and conformer-level stereochemical accuracy by 0.60 pp. Removing Conditional ProductLite decreases COV-P by 1.16 pp and increases MAT-P by 0.018 Å. These trends support the inclusion of both components in the precision-oriented model, although neither improves every metric. Removing AttnRes slightly *increases* COV-F1 by 0.03 pp while increasing MAT-F1 by 0.004 Å; this comparison does not demonstrate a coverage benefit from AttnRes.

Explicit topology and stereochemical loss show distinct effects. Removing chirality loss decreases overall conformer-level stereochemical accuracy by 0.54 pp, with only a small change in COV-F1 (-0.10 pp). Jointly removing the explicit hop-2/3 classes has a modest effect on overall COV-F1 (-0.24 pp), but decreases six-mer COV-F1 by 2.23 pp and six-mer conformer-level stereochemical accuracy by 1.08 pp. Because radius-based and global communication remain available, this ablation tests the explicit representation of short-hop topology rather than the complete removal of interactions within these atom pairs.

Optimization alters the recall–precision trade-off. At the same stated learning rate of 8×10^{-4}, AdamW increases COV-R by 2.53 pp relative to SOAP, but decreases COV-P by 6.38 pp, increases MAT-P by 0.093 Å, and decreases COV-F1 by 2.06 pp. These results support the use of SOAP for the evaluated configuration, but do not establish superiority over independently tuned AdamW settings or distinguish convergence speed from the quality of the selected checkpoint.

The joint five-strip (AdamW) configuration retains chirality and formal-charge conditioning, covalent and radius edges, and the SO(3)-equivariant cross-product term. Relative to the full SOAP parent model, its combined ("Together") COV-F1 decreases from 76.05% to 71.49%, COV-P decreases by 10.31 pp, and conformer-level stereochemical accuracy decreases by 1.73 pp. Six-mer COV-F1 decreases from 57.94% to 33.44%, whereas four-mer recall improves. Because the effects of individual ablations are relatively modest, the five-strip configuration provides a complementary joint ablation that highlights their cumulative impact, particularly for the more challenging six-mer systems.

D Evaluation Definitions and Aggregation

Heavy-atom matching. Hydrogens are removed before full-molecule RMSD evaluation. RDKit's best-alignment procedure considers chemically equivalent atom mappings to avoid artificial positional errors from exchanging equivalent atoms. Each mapping is aligned using translation and proper rotation. For each molecule m, the minimum RMSD is computed for every reference conformer over all generated conformers (recall) and for every generated conformer over all reference conformers (precision). Coverage uses the strict threshold RMSD < 0.75 Å.

Molecule-level means and harmonic summaries. For an individual molecule m, let $a_m \; and \; b_m$ denote paired recall and precision metrics, respectively, with overbars representing molecule-level averages. For a split s containing M_s molecules, the mean metric is defined as $\overline{a}_s = M_s^{-1} \sum_{m \in s} a_m$. Subscripts 4, 5, and 6 denote peptide-length subsets, yielding the combined test-set mean $\overline{a} = (533\overline{a}_4 + 347\overline{a}_5 + 120\overline{a}_6)\,/1000$ (and analogously for the precision mean $\overline{b}$). For evaluated models, the harmonic mean $F1\left(\overline{a}, \overline{b}\right) = 2\overline{a}\overline{b}/\left(\overline{a} + \overline{b}\right)$ is computed after aggregation, with zero assigned if the denominator is zero. This post-aggregation score differs from the mean of individual harmonic means, $M^{-1} \sum_m F1\left(a_m, b_m\right)$ across M molecules; thus, the combined F1 is not a weighted average of length-specific F1 values. Coverage is expressed in percent and matching in Å, while literature-baseline F1 values are retained as originally reported.

Stereochemical reference and denominators.

The metric evaluates source-specified RDKit tetrahedral CW/CCW centers with an ordered neighbor triple. For each usable center, the reference orientation is the modal nonzero scalar-triple-product sign over reference conformers. The scorer thresholds the raw, unnormalized scalar triple product at $\pm 10^{-7}$ Å^3 for coordinates in Å, unlike the normalized-volume training loss. The reference sign is whichever of the positive and negative signs occurs more often; equal counts, including all-zero references, make the center unusable. A generated zero/near-planar sign fails for a usable center. Unusable centers are excluded. This scoring reference is distinct from the training loss, which takes the sign of the currently sampled training conformer.

Let C_{mg} denote usable centers in generated conformer g, and b_{mgc} the orientation-match indicator. The two metrics are

$$Acc_{conf} = 100 \frac{\sum_{m,g:\,|C_{mg}|>0} \mathbf{1}\left[\forall c \in C_{mg},\ b_{mgc} = 1\right]}{\sum_{m,g} \mathbf{1}\left[|C_{mg}| > 0\right]}, \tag{41}$$

$$Acc_{center} = 100 \frac{\sum_{m,g} \sum_{c \in C_{mg}} b_{mgc}}{\sum_{m,g} |C_{mg}|}. \tag{42}$$

Here m, g, c index molecules, generated conformers, and usable centers; C_{mg} is the usable-center set and b_{mgc} is 1 for a reference-orientation match and 0 otherwise. The percentage accuracies Acc_{conf} and Acc_{center} pool fully matching conformers with at least one usable center and matching center instances, respectively.

These are generated-conformer and center-count pools, respectively, not molecule means. The native conformer denominator excludes records without usable centers. The checked canonical summaries have a usable center for every generated record, making pooling by total generated count equivalent for this panel.

E Dataset and Comparator Provenance

E.1 Dataset reconstruction and sample counts

The training and validation sets were reconstructed by deterministically dividing the original RINGER training partition using a length-stratified 90/10 split with seed 6489. The test set is the original RINGER/CREMP held-out set. No SMILES identifiers are shared among the training, validation, and test manifests. This establishes separation of the recorded molecular identities. Table 2 lists the molecule and conformer counts.

Table 2: Dataset partition sizes. The original training partition was stratified by peptide length and split with seed 6489; the original test partition remained held out.

Partition	4-mer	5-mer	6-mer	All molecules	Conformers
Train	15578	11967	4133	31678	948125
Validation	1731	1330	459	3520	105302
Test	533	347	120	1000	877898

The training and validation sets retain at most the 30 lowest-energy conformers per molecule; evaluation uses all 877,898 retained test references. Generating exactly $2K$ conformers per molecule yields 1,755,796 generated conformers in total. Molecular identities do not overlap across partitions.

E.2 Literature baseline rows

The six dagger-marked rows in Figure 3—RDKit, OMEGA, GeoDiff-Macro, DMCG-Macro, TorDiff-Macro, and RINGER—are taken from the RINGER benchmark (Grambow et al., 2024b), rather than newly generated in this study. Table 3 retains the published values.

The nominal $2K$ sampling budget described in the literature does not establish that classical methods retain exactly $2K$ final conformers: embedding failures, energy optimization, and RMSD pruning can affect the final count. In particular, the absence of force-field relaxation in CycPepFlow should not be attributed to the RDKit or OMEGA pipelines. These literature baselines provide context, but do not constitute an exactly budget-matched controlled comparison.

E.3 CREMP retraining and output protocols

The locally retrained LoQI, ET-Flow, and four DiTMC variants each completed 200 epochs. LoQI and ET-Flow use a global batch size of 64, whereas DiTMC uses 128. Their molecule-level loaders

sample one retained training conformer per molecule access or epoch. By contrast, MCF-S completed 750,000 optimizer steps with a global batch size of 128 over 948,125 retained conformer-level training examples. Its reported result uses the 300,000-step EMA checkpoint, selected using a validation diagnostic that evaluates 20 conformers for each of 20 validation molecules. Thus, these models share the retained top-30 conformer pool, but example exposure, validation usage, and training budgets are model specific rather than identical to the original models' settings.

- **LoQI-Macro** is the CREMP adaptation evaluated in this study. The original LoQI datasets and evaluations are not used as substitutes for the CREMP results reported here.
- **ETFlow-Macro** uses the ET-Flow(SO3) model retrained on CREMP, with epoch 191 selected for evaluation (Hassan et al., 2024).
- **MCF-S-Macro** uses the validation-selected 300,000-step EMA checkpoint rather than the terminal 750,000-step checkpoint, which showed degraded performance. Checkpoint selection used the recorded validation diagnostic. MCF's training and validation procedures differ from CycPepFlow's molecule-sampling procedure (Wang et al., 2024).
- The DiTMC aPE/rPE B/L results are evaluated using saved, raw generated coordinates. Every variant–length group retains exactly $2K$ conformers per molecule, with no parity correction, filtering, or force-field minimization in the raw-scoring protocol (Frank et al., 2025).
- **CycPepFlow** B/L and APG-B/L use the architectures and original-run settings described in Appendix A. All four share the same global training batch size. CycPepFlow-B with SOAP is the parent model for the ablations.

F Complete Evaluation Tables

Tables 3–10 report all 526 numerical metric entries in Figures 3–4 and Figure 9, comprising 17 overall rows, 21 length-specific rows, and 32 ablation-by-split rows. Bold type identifies the best value within each displayed comparison; for evaluations performed in this study, values are compared before rounding. No second-best highlighting is used. Coverage and stereochemical accuracy are percentages, and matching is in Å. The accompanying machine-readable files retain the original numerical precision.

Table 3: Complete results corresponding to Figure 3: combined ("Together") evaluation of 4-, 5-, and 6-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓
RDKit†	41.10	7.50	12.70	0.853	1.419	1.065
OMEGA†	36.20	6.40	10.90	0.900	1.433	1.106
GeoDiff-Macro†	25.60	2.70	4.90	0.989	1.753	1.265
DMCG-Macro†	**78.30**	41.90	54.60	0.534	0.942	0.682
TorDiff-Macro†	44.40	8.30	14.00	0.820	1.399	1.034
RINGER†	62.30	76.20	68.60	0.684	0.525	0.594
LoQI-Macro	77.85	50.68	61.39	0.524	0.859	0.651
ETFlow-Macro	63.53	38.77	48.15	0.649	1.051	0.802
MCF-S-Macro	60.41	39.77	47.96	0.691	1.055	0.835
DiTMC rPE-B-Macro	74.79	71.86	73.30	0.537	0.534	0.535
DiTMC rPE-L-Macro	72.94	76.36	74.61	0.549	0.464	0.503
DiTMC aPE-B-Macro	78.19	66.70	71.99	0.506	0.611	0.554
DiTMC aPE-L-Macro	77.85	73.02	75.36	**0.495**	0.514	0.504
CycPepFlow-B	74.95	77.18	76.05	0.531	0.461	0.494
CycPepFlow-L	74.16	78.21	**76.13**	0.539	0.441	0.485
CycPepFlow APG-B	73.81	77.99	75.85	0.543	0.449	0.492
CycPepFlow APG-L	72.65	**79.78**	76.05	0.555	**0.419**	**0.477**

COV is reported as a percentage; MAT is in Å. Bold type indicates the best value for each metric among the displayed methods. For models evaluated in this study, rankings use unrounded values, and F1 is the harmonic mean of molecule-averaged recall and precision, not the mean of molecule-level F1 values. † marks the six literature-transcribed rows in Figure 3; their F1 values are retained as reported rather than recalculated from rounded recall and precision. The remaining Macro rows are CREMP adaptations evaluated in this study, not the original papers' benchmark results. DiTMC uses raw coordinates without parity correction.

Table 4: Complete length-stratified results corresponding to Figure 4: 4-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
LoQI-Macro	**84.89**	64.62	73.38	**0.438**	0.641	0.520	98.47	99.65
ETFlow-Macro	72.55	51.63	60.33	0.552	0.841	0.667	48.91	79.88
MCF-S-Macro	67.86	51.53	58.58	0.613	0.874	0.720	45.87	77.74
DiTMC rPE-L-Macro	76.02	82.59	79.17	0.489	0.351	0.409	94.76	97.38
DiTMC aPE-L-Macro	80.10	80.60	**80.35**	0.446	0.385	0.413	96.04	98.06
CycPepFlow-L	75.45	84.78	79.84	0.503	0.326	0.395	**99.77**	**99.94**
CycPepFlow APG-L	73.79	**86.06**	79.45	0.525	**0.307**	**0.388**	99.74	99.90

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type indicates the best value for each metric within the displayed split, determined before rounding. F1 is the harmonic mean of unrounded molecule-averaged recall and precision, not the mean of molecule-level F1 values. "Conf. acc." and "Center acc." denote conformer-level and stereocenter-level agreement with source-specified tetrahedral orientations; neither assesses general geometric validity. All seven rows are CREMP benchmark results obtained in this study. DiTMC uses raw coordinates without parity correction. Figure 4 excludes B variants and RINGER.

Table 5: Complete length-stratified results corresponding to Figure 4: 5-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
LoQI-Macro	75.04	42.97	54.65	0.572	0.927	0.707	95.92	99.26
ETFlow-Macro	58.11	29.03	38.72	0.714	1.156	0.882	30.49	75.17
MCF-S-Macro	55.82	30.96	39.83	0.743	1.146	0.902	27.59	73.74
DiTMC rPE-L-Macro	70.22	75.70	72.86	0.586	0.470	0.522	92.39	97.35
DiTMC aPE-L-Macro	**76.14**	71.10	73.54	**0.530**	0.545	0.537	93.47	97.43
CycPepFlow-L	73.00	78.03	**75.43**	0.562	0.449	0.499	99.61	99.91
CycPepFlow APG-L	71.20	**79.08**	74.94	0.575	**0.434**	**0.495**	**99.78**	**99.94**

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type indicates the best value for each metric in this split, determined before rounding. Metric definitions, aggregation, and raw-output evaluation protocols are the same as in Table 4.

Table 6: Complete length-stratified results corresponding to Figure 4: 6-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
LoQI-Macro	54.66	11.07	18.41	0.764	1.627	1.040	92.86	98.89
ETFlow-Macro	39.09	9.77	15.63	0.893	1.680	1.166	25.49	74.10
MCF-S-Macro	40.57	12.97	19.65	0.890	1.598	1.143	24.53	73.86

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
DiTMC rPE-L-Macro	67.16	50.59	57.71	0.705	0.948	0.808	91.30	97.87
DiTMC aPE-L-Macro	**72.76**	44.95	55.57	**0.613**	0.999	0.760	93.14	98.42
CycPepFlow-L	71.75	49.61	58.66	0.635	0.933	0.755	99.44	99.91
CycPepFlow APG-L	71.78	**53.96**	**61.61**	0.635	**0.867**	**0.733**	**99.71**	**99.95**

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type indicates the best value for each metric in this split, determined before rounding. Metric definitions, aggregation, and raw-output evaluation protocols are the same as in Table 4.

Table 7: Complete component and optimizer ablation results corresponding to Figure 9: 4-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
Original ProductLite h192 SOAP LR 8e-4	76.41	84.19	80.11	0.492	0.340	0.402	**99.78**	**99.94**
No AttnRes	77.27	84.12	80.55	0.487	0.345	0.404	99.67	99.90
No gated global attention	78.52	83.35	**80.86**	0.469	0.359	0.407	99.54	99.87
No chirality loss	76.58	**84.47**	80.33	0.489	**0.337**	**0.399**	99.47	99.85
No MACEProductLite	76.96	83.31	80.01	0.485	0.357	0.411	99.60	99.86
Original ProductLite h192 AdamW LR 8e-4	80.33	80.54	80.43	0.459	0.406	0.431	99.43	99.80
No explicit hop2 / hop3 (joint)	76.96	84.22	80.43	0.487	0.343	0.402	99.62	99.90
Five-strip (AdamW)	**81.26**	79.76	80.50	**0.459**	0.429	0.443	99.14	99.74

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type marks the best value per metric and split, determined before rounding. F1 is computed from unrounded molecule-averaged recall and precision, not averaged over molecule-level F1 values. Accuracy measures source-specified tetrahedral agreement, not general geometric validity. The parent is CycPepFlow-B (Original ProductLite h192, SOAP). The joint hop ablation removes explicit hop-2/3 classes; covalent and dynamic radius edges remain, and affected pairs may still appear as radius neighbors. The no-loss ablation retains chirality node features. Five-strip (AdamW) removes AttnRes, gated global attention, ProductLite, chirality loss, and explicit hop-2/3 classes, and replaces SOAP with AdamW; chirality and formal-charge inputs and covalent and radius edges remain. "Together" pools COV/MAT by molecule counts and accuracy by generated-conformer or stereocenter counts.

Table 8: Complete component and optimizer ablation results corresponding to Figure 9: 5-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
Original ProductLite h192 SOAP LR 8e-4	73.74	**76.35**	**75.02**	0.557	**0.479**	**0.515**	**99.58**	**99.90**
No AttnRes	74.10	75.77	74.92	0.553	0.488	0.519	99.50	99.86
No gated global attention	74.70	74.02	74.36	0.550	0.513	0.531	98.89	99.62
No chirality loss	73.49	76.06	74.76	0.558	0.483	0.518	98.88	99.65
No MACEProductLite	73.82	75.23	74.52	0.558	0.493	0.524	99.49	99.82

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
Original ProductLite h192 AdamW LR 8e-4	**76.16**	68.04	71.87	**0.540**	0.590	0.564	98.84	99.64
No explicit hop2 / hop3 (joint)	73.11	76.17	74.61	0.562	0.483	0.520	99.25	99.86
Five-strip (AdamW)	75.76	62.20	68.31	0.554	0.675	0.609	97.66	99.41

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type marks the best value for each metric in this split, determined before rounding. The parent model, ablation definitions, and aggregation conventions are the same as in Table 7.

Table 9: Complete component and optimizer ablation results corresponding to Figure 9: 6-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
Original ProductLite h192 SOAP LR 8e-4	**71.95**	**48.49**	**57.94**	**0.629**	**0.949**	**0.757**	**99.37**	**99.89**
No AttnRes	71.00	46.89	56.48	0.643	0.966	0.772	99.26	99.88
No gated global attention	69.91	41.87	52.37	0.647	1.031	0.795	98.05	99.66
No chirality loss	71.66	46.67	56.53	0.632	0.974	0.767	98.67	99.78
No MACEProductLite	70.80	45.97	55.75	0.635	0.979	0.771	99.31	99.88
Original ProductLite h192 AdamW LR 8e-4	68.66	35.56	46.85	0.655	1.112	0.825	98.62	99.77
No explicit hop2 / hop3 (joint)	69.75	46.37	55.71	0.645	0.984	0.779	98.29	99.73
Five-strip (AdamW)	59.88	23.19	33.44	0.727	1.278	0.927	95.15	99.12

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type marks the best value for each metric in this split, determined before rounding. The parent model, ablation definitions, and aggregation conventions are the same as in Table 7.

Table 10: Complete component and optimizer ablation results corresponding to Figure 9: combined ("Together") evaluation of 4-, 5-, and 6-mers.

Method	COV-R ↑	COV-P ↑	COV-F1 ↑	MAT-R ↓	MAT-P ↓	MAT-F1 ↓	Conf. acc. ↑	Center acc. ↑
Original ProductLite h192 SOAP LR 8e-4	74.95	**77.18**	76.05	0.531	**0.461**	**0.494**	**99.63**	**99.91**
No AttnRes	75.42	76.75	**76.08**	0.529	0.469	0.497	99.53	99.88
No gated global attention	76.16	75.14	75.65	0.519	0.493	0.506	99.03	99.72
No chirality loss	74.92	77.02	75.95	0.530	0.464	0.495	99.09	99.75
No MACEProductLite	75.13	76.03	75.58	0.529	0.479	0.503	99.51	99.85
Original ProductLite h192 AdamW LR 8e-4	**77.48**	70.81	73.99	**0.511**	0.555	0.532	99.05	99.72
No explicit hop2 / hop3 (joint)	74.76	76.89	75.81	0.532	0.468	0.498	99.26	99.85
Five-strip (AdamW)	76.78	66.88	71.49	0.524	0.616	0.566	97.90	99.47

COV and stereochemical accuracy are percentages; MAT is in Å. Bold type marks the best value for each metric, determined before rounding. The parent model and ablation definitions are the same as

in Table 7. Combined recall and precision metrics are averaged over molecules before computing F1; stereochemical accuracies use generated-conformer or stereocenter-count pooling.

Table 11: Aggregate ("Together") differences from the full CycPepFlow-B parent (Original ProductLite h192, SOAP) in Figure 9.

Variant minus parent	**ΔCOV-R**	**ΔCOV-P**	**ΔCOV-F1**	**ΔMAT-R**	**ΔMAT-P**	**ΔMAT-F1**	**ΔConf. acc.**	**Δ Center acc.**
No AttnRes	+0.4685	-0.4323	+0.0287	-0.0023	+0.0081	+0.0036	-0.0985	-0.0356
No gated global attention	+1.2120	-2.0484	-0.4053	-0.0125	+0.0319	+0.0119	-0.5980	-0.1989
No chirality loss	-0.0333	-0.1653	-0.0974	-0.0007	+0.0032	+0.0015	-0.5395	-0.1661
No MACEProductLite	+0.1828	-1.1583	-0.4739	-0.0024	+0.0177	+0.0089	-0.1199	-0.0666
AdamW (full architecture)	+2.5318	-6.3794	-2.0577	-0.0200	+0.0933	+0.0382	-0.5769	-0.1904
No explicit hop2/3 (joint)	-0.1908	-0.2996	-0.2437	+0.0012	+0.0071	+0.0046	-0.3746	-0.0681
Five-strip (AdamW)	+1.8340	-10.3067	-4.5615	-0.0071	+0.1548	+0.0726	-1.7341	-0.4446

Entries are variant-minus-parent differences. Coverage and stereochemical accuracy differences are in percentage points; matching differences are in Å. Positive coverage or accuracy differences and negative matching differences favor the variant. These are descriptive single-run comparisons, not statistical significance tests. Best-value highlighting is not applied to this difference-only table. Removing explicit hop-2/3 classes leaves the dynamic radius path intact.