

# RINGER: CONFORMER ENSEMBLE GENERATION OF MACROCYCLIC PEPTIDES WITH SEQUENCE-CONDITIONED INTERNAL COORDINATE DIFFUSION

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Macrocyclic peptides are an emerging therapeutic modality, yet computational approaches for accurately sampling their diverse 3D ensembles remain challenging due to their conformational diversity and geometric constraints. Here, we introduce RINGER, a diffusion-based transformer model based on redundant internal coordinates **that generates three-dimensional conformational ensembles of macrocyclic peptides from their 2D representations**. RINGER provides fast backbone- and side-chain sampling while respecting key structural invariances of cyclic peptides. Through extensive benchmarking and analysis against gold-standard conformer ensembles of cyclic peptides generated with metadynamics, we demonstrate how RINGER generates both high-quality and diverse geometries at a fraction of the computational cost. Our work lays the foundation for improved sampling of cyclic geometries and the development of geometric learning methods for peptides.

## 1 INTRODUCTION

Macrocyclic peptides are an important therapeutic modality in modern drug discovery that occupy a unique chemical and pharmacological space between small and large molecules (Driggers et al., 2008; Muttenthaler et al., 2021; Vinogradov et al., 2019). These cyclic peptides exhibit improved structural rigidity and metabolic stability compared to their linear counterparts (Craig et al., 2013), yet retain key conformational flexibility and diversity to bind shallow protein interfaces (Villar et al., 2014). However, computational approaches for modeling their structural ensembles remain limited compared to small molecules and proteins in terms of computational speed, accuracy (sample quality), and conformational diversity (Poongavanam et al., 2018). Critically, scalable and accurate tools are necessary to enable rational design of macrocyclic drugs; access to these tools can significantly impact optimization of key properties including binding affinity (Alogheli et al., 2017; Garcia Jimenez et al., 2023), permeability (Leung et al., 2016; Rezai et al., 2006; Bhardwaj et al., 2022), and oral bioavailability (Nielsen et al., 2017).

Several key challenges hinder fast and effective macrocycle conformer generation: 1) Macrocyclic peptides exhibit diverse molecular structures and chemical modifications, including varying ring size, stereochemistry, *N*-methylation, and more (Kamenik et al., 2018). Their structural diversity, along with the increased number of rotatable bonds, results in a vast conformational space that is considerably more expensive to sample computationally. 2) Macrocycles are subject to complex non-linear constraints due to ring closure. The atomic positions, angles, and dihedrals of the macrocycle backbone are highly interdependent, and additional complex intramolecular interactions make this process inherently difficult to model (Watts et al., 2014). 3) Experimental X-ray and NMR structures for macrocycles are lacking ( $\sim 10^3$ ) in comparison to small molecules ( $\sim 10^6$  in the Cambridge Structural Database (Groom et al., 2016)) and proteins ( $\sim 10^5$  in the Protein Data Bank (Berman et al., 2000)). The scarcity of available experimental data has made it difficult to integrate observational data to improve structural predictions or train machine learning-based approaches. Together, the vast conformational space combined with limited data make modeling and sampling of macrocycles not only conceptually challenging, but technically challenging due to computational cost. **Existing approaches cannot accurately generate diverse and accurate conformational ensembles at a scale necessary to accelerate and enable macrocycle design.**

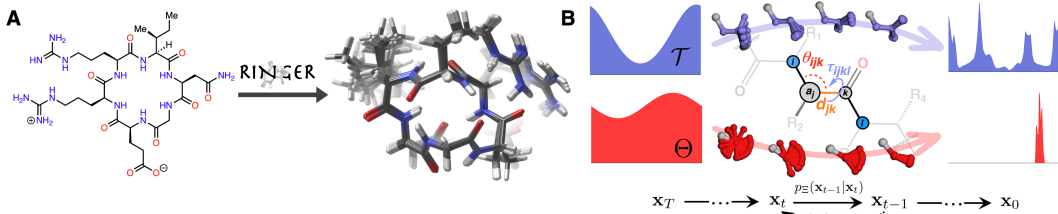


Figure 1: Overview of RINGER for macrocycle conformer generation. **A.** Given a 2D representation of a macrocyclic peptide, RINGER generates an accurate and diverse 3D conformational ensemble. **B.** An illustration of the diffusion process learning to recover the time  $t = 0$  bond angle (red) and torsional (blue) distributions from time,  $t = T$ .

To address these limitations, we introduce RINGER (RINGER Generates Ensembles of Rings), a deep learning model designed specifically for sequence-conditioned macrocycle conformer generation (Figure 1) that efficiently samples realistic angles and torsions (i.e., internal coordinates) for macrocyclic peptides. RINGER merges a transformer architecture that naturally captures the physical equivariants and invariants of macrocyclic peptides with a discrete-time diffusion model to learn highly-coupled distributions over internal coordinates. We demonstrate how RINGER simultaneously achieves excellent performance in sample quality over angular and torsional profiles while maintaining excellent RMSDs relative to gold-standard conformer ensembles generated with the Conformer-Rotamer Ensemble Sampling Tool (CREST) (Pracht et al., 2020).

We summarize our contributions as follows:

- We propose a new framework, RINGER for all-atom conformer generation of macrocycles based on efficiently encoding their geometry using redundant internal coordinates. Our model naturally handles the cyclic nature of macrocycles and chiral side chains with both L- and D-amino acids, and we propose a simple solution to satisfy ring-closure constraints.
- We benchmark RINGER extensively against state-of-the-art physics- and machine learning-based algorithms to demonstrate how our approach better captures complex and highly-coupled geometric distributions. **RINGER outperforms state-of-the-art algorithms that predict 3D conformer ensembles from their 2D structures for macrocyclic peptides.**
- Our work highlights the advantages of a machine learning-based approach for conformer generation in quality, diversity, and speed. We show how RINGER generates diverse and complete conformational ensembles with excellent sample quality that nearly match the gold-standard ensembles generated with expensive metadynamic simulations.

## 2 BACKGROUND AND RELATED WORK

Our work builds on small-molecule conformer generation and protein structure modeling to create a framework for macrocycle conformers. Below, we briefly summarize related work.

**Physics and Heuristic-based Conformer Generation for Macrocycles** Physics-based and heuristic-based algorithms remain the state of the art for macrocycles and have required special considerations compared to drug-like small molecules due to ring-closing constraints. The open-source cheminformatics library RDKit leverages distance geometry algorithms for small-molecule conformer generation (ETKDG) (Riniker & Landrum, 2015), with improved heuristic bounds for macrocycles (ETKDGv3) (Wang et al., 2020; 2022). Similarly, commercial software such as OpenEye OMEGA (Hawkins et al., 2010; Hawkins & Nicholls, 2012) in `macrocycle` mode uses a distance geometry algorithm based on 4D coordinate initialization to provide diverse conformers (Spellmeyer et al., 1997), as their torsion-driving approach is incompatible with ring closure.

Similarly, low-mode (Kolossváry & Guida, 1996; 1999) or Monte Carlo (Chang et al., 1989) search methods combined with molecular dynamics have been found to be effective at sampling macrocycle conformations, particularly when combined with force field optimizations as demonstrated in Schrödinger’s MacroModel (Watts et al., 2014) and Prime MCS (Sindhikara et al., 2017). These

approaches have been tuned with expert knowledge and torsional libraries to maximize agreement with observed experimental structures. The open-source CREST package (Pracht et al., 2020) leverages iterative metadynamics with a genetic structure-crossing algorithm (iMTD-GC) to explore new geometries, and can be considered a gold-standard for generating diverse ensembles of drug-like molecules. In this work, we use the recently-published CREMP (Grambow et al., 2023) dataset, containing high-quality, CREST-generated ensembles, representing over 31 million macrocycle geometries (see Section 4.1 and Appendix B for more details).

One key limitation of these approaches is high computational cost and difficulty in scaling; in general, conformer generation is  $10^3 - 10^5 \times$  more computationally expensive compared to a drug-like small molecule due to the increased number of rotatable bonds and their ring-closing constraints (e.g., generating a conformational ensemble of a macrocyclic hexapeptide with CREST requires an average of 14 hours (Grambow et al., 2023)). These approaches become increasingly challenging when kinetic or molecular dynamics approaches are used with explicit solvation (Damjanovic et al., 2021; Linker et al., 2023).

**Generative Approaches for Small Molecule Conformer Ensembles** Recent work with deep generative models has focused on improved sampling of the conformational landscape of small molecules. For example, Mansimov et al. (2019) propose a conditional graph variational autoencoder (CGVAE) approach for molecular geometry generation. Simm & Hernandez-Lobato (2020) report conditional generation of molecular geometries based on distance geometry. Xu et al. (2021) leverage normalizing flows and energy-based modeling to help capture the multimodal nature and complex dependencies of small molecule space. More recently, Xu et al. (2022) report GeoDiff, an equivariant diffusion-based model that operates on Cartesian point clouds. Although GeoDiff provides strong results, sampling is costly and requires 5,000 time steps. Recently, Zhu et al. (2022) report DMCG, which leverages a variational autoencoder to directly predict coordinates while maintaining invariance to rototranslation and permutation of symmetric atoms.

Recent reports have also drawn inspiration from physics-based conformer generation to leverage the rigid-rotor hypothesis, which treats bond distances and angles as fixed, and torsional angles of rotatable bonds are independently sampled, assuming little or no interdependence between torsions (Schärfer et al., 2013). These include GeoMol (Ganea et al., 2021), an SE(3)-invariant machine learning model for small molecule conformer generation that leverages graph neural networks, and EquiBind (Stärk et al., 2022) which performs conditional generation on protein structure. Recently, Jing et al. (2022) report Torsional Diffusion, a diffusion model that operates on the torsional space via an extrinsic-to-intrinsic score model to provide strong benchmarks on the GEOM dataset (Axelrod & Gómez-Bombarelli, 2022). Importantly, these methods do not address the challenge of highly-coupled torsions within cyclic systems and either propose complex ring-averaging processes (Ganea et al., 2021) or disregard sampling of cyclic structures all together (Jing et al., 2022).

**Protein Structure Prediction and Diffusion** Significant progress has been made recently in protein structure prediction with the advent of methods such as AlphaFold2 (Jumper et al., 2021) and RoseTTAFold (Baek et al., 2021). However, structure prediction methods have predominantly focused on deterministic maps to static output structures rather than on sampling diverse structure ensembles. Recently, several papers have developed diffusion-based approaches for protein generation based on Euclidean diffusion over Cartesian coordinates (Anand & Achim, 2022; Yim et al., 2023) or backbones as in FoldingDiff (Wu et al., 2022), with an emphasis on structural design. FoldingDiff parameterizes structures over internal backbone angles and torsions and relies on the natural extension reference frame (NeRF) (Parsons et al., 2005) to perform linear reconstructions. However, as we demonstrate below, naive linear transformations fail to address the ring constraints for macrocycles. Moreover, FoldingDiff focuses on *unconditional* generation of protein backbones and is hence not suitable for all-atom conformer ensemble generation. Below, we focus on the challenging problem of conditional generation of constrained geometries.

**Machine Learning Approaches for Macrocycle Conformer Ensemble Generation** Despite the many approaches focused on small molecules and protein structure generation, there are few efforts in macrocycle structure prediction. Most notably, Miao et al. (2021) recently disclosed StrEAMM for learning on molecular dynamics of cyclic peptides using explicit solvation. StrEAMM is a linear model that predicts local backbone geometries and their respective 1,2- and 1,3-residue interactions to provide excellent ensemble estimates of homodetic hexapeptides, but does not generate explicit all-atom conformers. Additionally, the model is not naturally inductive and is not natively extensible

to other macrocycle ring sizes and residues. Fishman et al. (2023) recently developed a more general framework for diffusion models on manifolds defined via a set of inequality constraints. However, they only investigate the conformational ensemble of a single cyclic peptide as a proof-of-concept using a reduced  $\alpha$ -carbon representation.

### 3 RINGER: PROBLEM STATEMENT AND METHODS

#### 3.1 PROBLEM DEFINITION: CONDITIONAL MACROCYCLE CONFORMER GENERATION

The core objective of our work is to model the distribution of conformers for a macrocyclic peptide with a given amino acid sequence. Given a peptide macrocycle graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V}$  is the set of nodes (atoms) and  $\mathcal{E}$  is the set of edges (bonds), and  $n = |\mathcal{V}|$  our goal is to learn a distribution over the possible conformers. Let  $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$  be the set of conformers, where each conformer  $c_k \in \mathcal{C}$  represents a unique spatial arrangement of the atoms  $\mathcal{V}$ . Our task is to learn the distribution  $p(\mathcal{C} | \mathcal{G})$ , which represents the probability over the conformer ensemble  $\mathcal{C}$  given a molecular graph  $\mathcal{G}$ . Learning and sampling from this complex distribution is inherently challenging for most molecules, and is further complicated in macrocycles due to the highly-coupled nature of ring atoms. A perturbation to one part of the ring generally perturbs the others. Consequently, any model must account for the interdependence between atoms due to the cyclic constraints.

Given this problem, a good generative model ideally satisfies a few key properties: 1) Naturally encodes the physical and structural aspects of macrocyclic peptides. For example, cyclic peptides with only standard peptide bonds (i.e., homodetic peptides) do not have a natural starting residue and hence exhibit cyclic shift invariance, e.g., cyclo-(R.I.N.G.E.R) is identical to cyclo-(I.N.G.E.R.R), where each amino acid is denoted by its one-letter code with “cyclo” indicating cyclization of the sequence. 2) Captures multimodal distributions and complex, higher-order interactions such as the strong coupling between atomic positions in the ring. 3) Samples high-quality and diverse conformations from  $p(\mathcal{C} | \mathcal{G})$  that faithfully capture realistic geometries while respecting the underlying conformer distribution.

#### 3.2 REPRESENTING MACROCYCLE GEOMETRY: REDUNDANT INTERNAL COORDINATES

Conformer geometries are defined by their set of Cartesian coordinates for each atomic position and can hence be modeled using SE(3)-equivariant models to learn complex distributions. However, Euclidean diffusion requires modeling the many degrees of freedom; and, in practice, can require many time steps to generate accurate geometries (Xu et al., 2022). Moreover, realistic conformations are highly sensitive to the precise interatomic distances, angles, and torsions—although this information is implicit in the Cartesian positions, explicitly integrating these quantities into a model can provide a strong inductive bias and accelerate learning (Gasteiger et al., 2020).

Borrowing from molecular geometry optimization (Peng et al., 1996), protein representation (Ramachandran & Sasisekharan, 1968; Dunbrack & Karplus, 1994; Parsons et al., 2005), and inverse kinematics (Han & Rudolph, 2006), we adopt redundant internal coordinates that represent conformer geometries through a set of bond distances, angles, and torsions (dihedral angles), i.e.,  $\mathcal{C} \equiv \{\mathcal{D}, \Theta, \mathcal{T}\}$ . In particular, this simplifies the learning task, as bond distances can be approximated as fixed distances with little loss in accuracy (Hawkins et al., 2010; Wu et al., 2022; Jing et al., 2022), and internal angles typically fit a narrow distribution. Importantly, these coordinates define an internal reference frame that readily encodes complex geometries including chirality. Moreover, this approach obviates the need for complex equivariant networks (Xu et al., 2022; Jing et al., 2022). Hence, our generative process can be reformulated as learning the distribution  $p(\{\Theta, \mathcal{T}\} | \mathcal{G}; \mathcal{D})$  using fixed bond distances for reconstruction back to Cartesians (Figure 1).

#### 3.3 DEEP PROBABILISTIC DIFFUSION MODELS FOR SAMPLING INTERNAL COORDINATES

**Denosing Probabilistic Models** Recent works on deep denoising probabilistic models have demonstrated excellent generative performance for complex multimodal data (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021), and have been successfully applied to both small molecules and proteins (Xu et al., 2022; Wu et al., 2022). We adapt a discrete-time diffusion process (Wu et al., 2022) that formulates the forward transition probability using a wrapped normal distribution,  $q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}_{\text{wrapped}}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I})$ , instead of a standard normal distribution (Jing et al., 2022), where  $\mathbf{x}_t$  represents the noised internal coordinates (bond angles and torsions) at time step  $t$ . We train a diffusion model,  $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t)$ , by training a neural architecture, described below,

to predict the noise present at a given time step (for full details, see Appendix C). During inference, we sample  $\mathbf{x}_T$  from a wrapped normal distribution and iteratively generate  $\mathbf{x}_0$  using  $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t)$ . The sampling process is further detailed in Appendix D.

**Encoder Architecture** Macrocycles exhibit extensive coupling of their residues due to torsional strain and intramolecular interactions such as hydrogen bonds. Here, we use self-attention (Vaswani et al., 2017; Devlin et al., 2019) to learn the complex interactions between atoms. Unlike standard sequence models for linear data, macrocycles exhibit cyclic symmetry with no canonical start position. Thus, we design a modified bidirectional, relative positional encoding (Shaw et al., 2018)  $\mathbf{p}_{ij}^K$  to reflect this physical invariance (see Appendix A for notation and Appendix N.4 for ablation study):

$$\mathbf{z}_i = \sum_{j=1}^n \alpha_{ij} (\mathbf{v}_j \mathbf{W}^V), \quad \text{where} \quad \alpha_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^n \exp e_{ik}} \quad (1)$$

$$e_{ij} = \frac{\mathbf{v}_i \mathbf{W}^Q (\mathbf{v}_j \mathbf{W}^K + \mathbf{p}_{ij}^K)^T}{\sqrt{d_z}} \quad \text{with} \quad \mathbf{p}_{ij}^K = \underbrace{\mathbf{W}_{(i-j) \bmod n}^D}_{\text{forward}} + \underbrace{\mathbf{W}_{(i-j) \bmod (-n)}^D}_{\text{backward}} \quad (2)$$

These cyclic relative position representations specify forward (N-to-C) and reverse (C-to-N) relationships between each atom in the macrocycle, and effectively encode the molecular graph in a self-attention module similar to Ying et al. (2021). The relative position of any neighboring atom is uniquely defined by its forward and reverse graph distances in the embedding lookup  $\mathbf{W}^D$ . For conditional generation, we perform a linear projection of the atom features  $\mathbf{a}_i$ , corresponding to *each macrocycle backbone atom and its side chain*, and a separate linear projection of the angles and torsions  $\mathbf{x}_i = \boldsymbol{\theta}_i \oplus \boldsymbol{\tau}_i$  and concatenate them as a single input to the transformer,  $\mathbf{v}_i = \mathbf{a}'_i \oplus \mathbf{x}'_i$ . Notably, our diffusion model only adds noise to the angular component,  $\mathbf{x}_i$ , corresponding to backbone and side-chain angles and torsions that define a complete atomic configuration. For unconditional backbone generation, atoms are only labeled with their backbone identity (nitrogen,  $\alpha$ -carbon, carbonyl-carbon) using an embedding that is added to the input, and side chains are not modeled. Model details are shown in Appendix E.

**Ring Closing: Back Conversion to Cartesian Ring Coordinates** Macrocycles with fixed bond distances contain three redundant torsional angles and two redundant bond angles. Whereas linear peptides and proteins can be readily converted into an arbitrary Cartesian reference frame through methods such as NeRF (Parsons et al., 2005), these redundancies prevent direct transformation to unique Cartesians for cyclic structures. Adopting a sequential reconstruction method such as NeRF accumulates small errors that result in inadequate ring closure for macrocycles.<sup>1</sup> Other studies have developed complex heuristics with coordinate averaging for ring smoothing (Ganea et al., 2021), yet these approaches can distort the predicted geometries. In practice, we demonstrate that an efficient post-processing step works well with minimal distortion: we treat this as a constrained optimization problem using the Sequential Least Squares Quadratic Programming (SLSQP) algorithm (Kraft, 1988) to ensure valid Cartesian coordinates while satisfying distance constraints:

$$\hat{\boldsymbol{\xi}} = \arg \min_{\boldsymbol{\xi}} \|\boldsymbol{\theta}(\boldsymbol{\xi}) - \hat{\boldsymbol{\theta}}\|^2 + \|w(\boldsymbol{\tau}(\boldsymbol{\xi}) - \hat{\boldsymbol{\tau}})\|^2 \quad \text{subject to:} \quad \mathbf{d}(\boldsymbol{\xi}) = \mathbf{d}_{\text{true}} \quad (3)$$

Here, we find the set of *ring* Cartesian coordinates,  $\hat{\boldsymbol{\xi}}$ , that minimize the squared error against the ring internal coordinates  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\tau}}$  sampled by the diffusion process while satisfying bond distance equality constraints using known ring bond distances,  $\mathbf{d}_{\text{true}}$ , from the training data. The torsion error,  $\boldsymbol{\tau}(\boldsymbol{\xi}) - \hat{\boldsymbol{\tau}}$ , is wrapped by  $w(\cdot)$  so that it remains in the  $[-\pi, \pi)$  range. Empirically, we demonstrate that this scheme recovers realistic macrocycles with high fidelity by evenly distributing the error across the entire macrocycle backbone (see Appendix F for additional details). Additionally, we reject a sample from RINGER if the ring torsion fingerprint deviation (Wang et al., 2020) before versus after optimization with equation 3 is larger than 0.01, which further improves the quality of generated ensembles.

<sup>1</sup>Although direct equality and inequality constraints over the diffusion process is a promising direction that could address this problem, we leave this for future work.

**Overall Generation Procedure** Our model represents macrocycles as cyclic sequences of backbone atoms with fixed bond lengths, where each atom is featurized with two ring internal coordinates and several side chain internal coordinates. We train a discrete-time diffusion model to learn a denoising process over the internal coordinates using a transformer architecture with an invariant cyclic positional encoding. Chemical features of the backbone atoms and molecular fingerprints of the attached side-chain atoms provide 2D macrocycle structure information for the sequence-conditioned models. At inference time, we sample from a wrapped Gaussian distribution to produce a set of angles and torsions. In the final post-processing step, macrocycle geometries with Cartesian coordinates are reconstructed through our constrained optimization using equation 3 for ring atoms and using NeRF (Parsons et al., 2005) for rotatable side-chain atoms. Hydrogens and non-rotatable side-chain groups like phenyl rings are not modeled directly and are generated using RDKit. A detailed flowchart illustrating the main steps of our method is shown in Appendix G.

## 4 EXPERIMENTS AND RESULTS

### 4.1 EXPERIMENTAL SETUP

**Dataset** We train and evaluate our approach on the recently published CREMP dataset (Grambow et al., 2023) that contains 36k homodetic macrocyclic peptides across varying ring sizes (4-mers, 5-mers, and 6-mers corresponding to 12-, 15-, and 18-membered backbone rings), side chains, amino-acid stereochemistry, and *N*-methylation. Each macrocycle in CREMP contains a conformational ensemble sampled with CREST (Pracht et al., 2020), a metadynamics algorithm with genetic crossing built on the semi-empirical tight-binding method GFN2-xTB (Bannwarth et al., 2019). We perform stratified random splitting on the data, with a training and validation set of 35,198 molecules (948,158 conformers using a maximum of 30 conformers per molecule), which we split into 90% training and 10% validation, and a final test set of 1,000 molecules corresponding to 877,898 distinct conformers (using *all* conformers per molecule within the 6 kcal/mol energy threshold defined by CREST). Additional dataset statistics are shown in Appendix B.

**Training & Sampling** All training is performed on the set of 35k peptides described above, using the 30 lowest-energy conformers per peptide. We train each model on a single NVIDIA A100 GPU for up to 1000 epochs until convergence using the Adam optimizer with 10 warmup epochs. Following work in small-molecule conformer generation (Ganea et al., 2021; Xu et al., 2022; Jing et al., 2022), we sample  $2K$  conformers for a macrocycle ensemble of  $K$  ground-truth conformers (median  $K = 656$ ) and assess them based on the evaluation criteria below. For full training and sampling details see Appendices C and D.

**Evaluation** For unconditional generation, we use Kullback-Leibler divergence to measure the difference in sample quality. For conditional generation, we evaluate the quality of our generated macrocycles using root-mean-squared-deviation (RMSD) between heavy-atom coordinates, similar to previous work on small-molecule conformer generation. We use several metrics including **Matching** and **Coverage** (Xu et al., 2021; Ganea et al., 2021; Jing et al., 2022), and for each we report recall, precision, and **F1-score**. We also report rRMSD, which is only evaluated on ring atoms. We note that although RMSD is widely used to assess conformer quality, its utility for comparing backbones is more limited, as sampled backbones with highly unrealistic or energetically unfavorable torsions can exhibit low rRMSD values. Therefore, we additionally report the ring torsion fingerprint deviation (rTFD) (Schulz-Gasch et al., 2012; Wang et al., 2020) to evaluate the quality of the torsional profiles. RMSD and rRMSD provide a measure of distance between two conformers based on a least-squares alignment of their respective atomic positions, while rTFD gives a normalized measure of matched torsion angles between backbone geometries. Appendix J defines the evaluation metrics in detail.

**Baselines** We provide benchmarks of our method against open-source and commercial toolkits RDKit ETKDGv3 (for macrocycles) (Wang et al., 2020), OMEGA Macrocycle Mode (OpenEye, 2022), GeoDiff (Xu et al., 2022), DMCG (Zhu et al., 2022), and Torsional Diffusion (Jing et al., 2022). For GeoDiff, DMCG, and Torsional Diffusion we report results trained on CREMP (GeoDiff-Macro, DMCG-Macro, and TorDiff-Macro), as the base models trained on small molecules provided poor performance. To better understand ring diversity, we additionally included a nearest neighbor model (1-NN) based on maximum graph similarity (2D) for ring-evaluation metrics only.

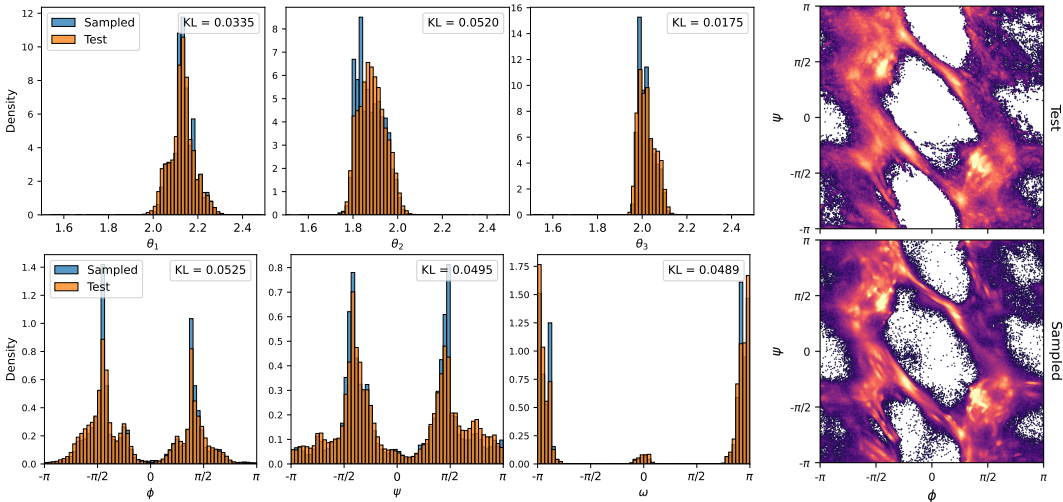


Figure 2: Unconditional conformer generation for a **backbone-only model**. Comparison of the bond angle and dihedral distributions from the held-out test set (orange) and in the *unconditionally* generated redundant internal coordinate samples (blue). The three top left plots correspond to the three bond angle types in each amino acid residue, the three bottom left plots show the three dihedral angles for each residue, and the right shows Ramachandran plots (colored logarithmically by density with high density regions shown in lighter colors). KL divergence is calculated as  $D_{\text{KL}}(\text{test} \parallel \text{sampled})$ .

#### 4.2 UNCONDITIONAL GENERATION OF MACROCYCLE BACKBONES

To understand whether this approach can learn the highly-coupled and underlying distribution of macrocycle conformations, we first trained RINGER on macrocycle backbones in the absence of any residue or side-chain features. From a design perspective, diverse backbone sampling alone can help drive inverse peptide design, where specific backbone geometries suggest important sequences. Figure 2 (and per-residue distributions in Appendix L) demonstrates how RINGER accurately replicates both angles and dihedrals with tight fidelity across all residue atoms. Furthermore, we generated Ramachandran plots (Ramachandran & Sasisekharan, 1968) alongside our withheld test set to visualize the conditional dependencies between residue torsions. Notably, RINGER recapitulates the critical modes of the distribution.

#### 4.3 CONDITIONAL GENERATION OF MACROCYCLE CONFORMATIONAL ENSEMBLES

We subsequently focused on the challenge of conditional generation to understand whether RINGER could effectively capture the complex steric and intramolecular effects that dictate macrocycle conformation. Whereas the unconditional model disregards side chains, we now condition generation on molecular features corresponding to each atom in the ring, including side-chain information, stereochemistry, and *N*-methylation (see Appendix E) to unambiguously predict all atomic internal coordinates. We evaluate RINGER both in the context of all-atom geometries (RMSD), but also evaluate on backbone-only ring geometries (rRMSD, rTFD) as they are critical for macrocycle design (Bhardwaj et al., 2022).

Comparison of RINGER RMSD, rRMSD, and rTFD ensemble metrics against the baselines are shown in Table 1. Here, recall quantifies the proportion of ground truth conformers that are recovered by the model, and precision quantifies the quality of the generated ensemble. We found that RDKit ETKDGv3 and OMEGA Macrocycle mode, both based on distance-geometry approaches, performed similarly across metrics and achieved moderate recall with limited precision. **Comparing to other deep learning approaches, we found that both GeoDiff and DMCG offer improved recall over heuristic baselines, with only DMCG generating improved precision and moderate F1-scores.** Moreover, the nearest neighbor baseline, which uses the backbone ensemble from the most similar training molecule (as measured by 2D similarity), provides only modest performance. This baseline demonstrates how side chain identity dictates 3D ensemble geometry, and that deep models cannot trivially memorize backbone geometries in this task. In contrast, RINGER achieves excellent performance across all metrics, providing strong recall and excellent precision over all test molecules, with the best F1-score.

Table 1: **Mean** performance metrics for sequence-conditioned generation of macrocycles. Coverage is evaluated at a threshold of 0.75 Å for all-atom RMSD, 0.1 Å for ring-only RMSD (rRMSD), and 0.05 for ring-only TFD (rTFD). *All test data conformers are used for evaluation.*

Method	RMSD – Recall		RMSD – Precision		RMSD - F1	
	COV (%) ↑	MAT ↓	COV (%) ↑	MAT ↓	COV (%) ↑	MAT ↓
RDKit (Wang et al., 2020)	41.1	0.853	7.5	1.419	12.7	1.065
OMEGA (OpenEye, 2022)	36.2	0.900	6.4	1.433	10.9	1.106
GeoDiff-Macro (Xu et al., 2022)	25.6	0.989	2.7	1.753	4.9	1.265
<b>DMCG-Macro</b> (Zhu et al., 2022)	<b>78.3</b>	<b>0.534</b>	41.9	0.942	54.6	0.682
<b>TorDiff-Macro</b> (Jing et al., 2022)	44.4	0.820	8.3	1.399	14.0	1.034
<b>RINGER</b>	62.3	0.684	<b>76.2</b>	<b>0.525</b>	<b>68.6</b>	<b>0.594</b>
	rRMSD – Recall		rRMSD – Precision		RMSD - F1	
1-NN (Seq. Sim.)	43.7	0.301	40.3	0.331	41.9	0.315
RDKit (Wang et al., 2020)	35.8	0.187	5.6	0.540	9.7	0.277
OMEGA (OpenEye, 2022)	32.2	0.186	3.7	0.557	6.6	0.279
GeoDiff-Macro (Xu et al., 2022)	50.8	0.151	6.4	0.592	11.4	0.240
<b>DMCG-Macro</b> (Zhu et al., 2022)	<b>77.6</b>	<b>0.076</b>	36.6	0.301	49.8	0.121
<b>TorDiff-Macro</b> (Jing et al., 2022)	35.8	0.187	5.6	0.540	9.7	0.277
<b>RINGER</b>	<b>76.3</b>	0.110	<b>81.6</b>	<b>0.108</b>	<b>78.9</b>	<b>0.109</b>
	rTFD – Recall		rTFD – Precision		RMSD - F1	
1-NN (Seq. Sim.)	53.1	0.111	48.6	0.122	50.7	0.116
RDKit (Wang et al., 2020)	52.9	0.059	9.4	0.215	15.9	0.093
OMEGA (OpenEye, 2022)	49.7	0.061	6.6	0.225	11.7	0.095
GeoDiff-Macro (Xu et al., 2022)	68.1	0.048	9.1	0.248	16.0	0.080
<b>DMCG-Macro</b> (Zhu et al., 2022)	<b>93.0</b>	<b>0.021</b>	48.5	0.110	63.8	<b>0.036</b>
<b>TorDiff-Macro</b> (Jing et al., 2022)	52.8	0.059	9.4	0.215	15.9	0.093
<b>RINGER</b>	83.8	0.035	<b>85.3</b>	<b>0.038</b>	<b>84.5</b>	<b>0.037</b>

As demonstrated through rRMSD and rTFD metrics, RINGER samples both diverse and high-quality backbone conformations.

#### 4.4 EVALUATING CONFORMER QUALITY WITH POST HOC OPTIMIZATION

The CREMP dataset (Grambow et al., 2023) used in this study leverages the semi-empirical extended tight binding method GFN2-xTB (Bannwarth et al., 2019) for geometry optimization. To enable a level comparison of sampling quality only, we reoptimized all generated samples from all methods with GFN2-xTB (Tables 11 and 12, Appendix O). Notably, we observed a dramatic improvement for xTB-optimized RDKit, OMEGA, GeoDiff-Macro, and TorDiff-Macro molecules, but only slight improvement in precision. In contrast, xTB-reoptimized samples from RINGER see only a minimal but consistent boost in performance, yet still outperform all other approaches. Further analysis of these xTB geometries revealed that RINGER-generated samples require smaller conformational adjustments to reach local minima (Figure 13, Appendix O), suggesting that our model can accurately learn the complex distributions directly from the training data. Overall, these results highlight two key strengths of our machine learning approach: 1) careful featurization enables generating high-quality conformations that closely match quantum chemical geometries, and 2) the diffusion scheme provides good sampling diversity.

#### 4.5 STRUCTURAL ANALYSIS OF GENERATED MACROCYCLES

Although RMSD and TFD evaluate performance quantitatively, analyzing individual ensembles elucidates the qualitative differences in conformer generation processes. Notably, the two distinct macrocycles in Figure 3 possess ensembles characterized by distinctly unique Ramachandran plots. Importantly, most ground truth conformer ensembles exhibit relatively tight distributions characterized by a distinctive set of  $\phi$ ,  $\psi$  angles (Reference, far left). Although all methods can identify relevant



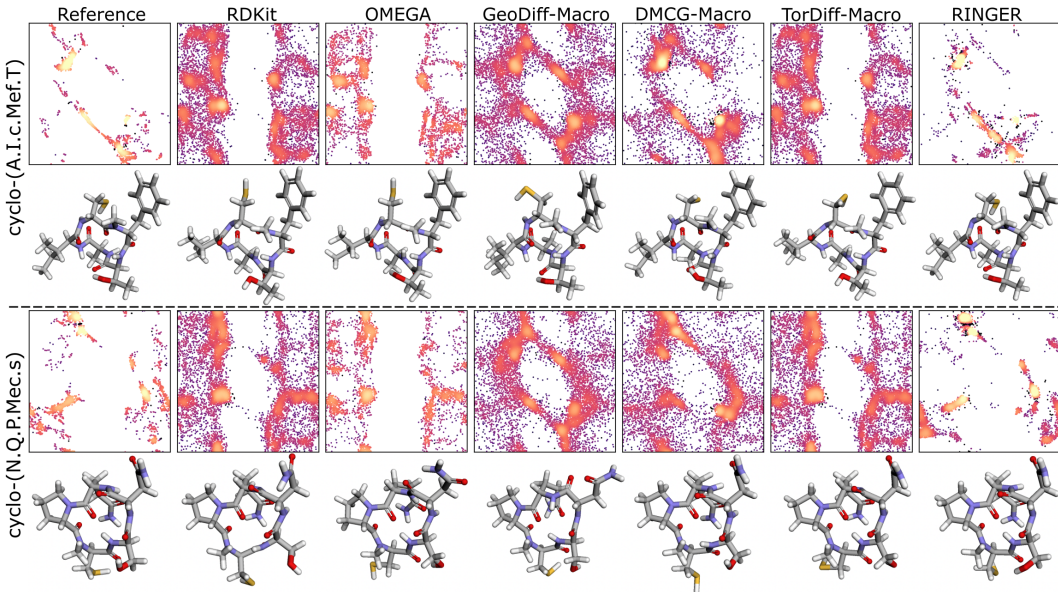


Figure 3: **Comparison of macrocycle conformational ensembles.** RINGER accurately generates ensembles as illustrated by Ramachandran plots for individual macrocycle ensembles, whereas other methods generate unrealistic backbone geometries. The individual 3D conformers illustrate the lowest-energy reference structure and the closest matching conformer (based on RMSD) from each method.

low-energy conformers, the overall sampling process generates unrealistic distributions for RDKit, OMEGA, and GeoDiff (note that TorDiff relies on RDKit for backbone sampling, and hence produces an identical plot). In contrast, RINGER recapitulates not only the ground state geometry with excellent accuracy, but better captures the entire ensemble distribution. These results demonstrate how RINGER achieves strong performance by providing high sample quality and diversity.

## 5 LIMITATIONS AND FUTURE DIRECTIONS

Our studies demonstrate the potential for diffusion-based models to overcome limitations in constrained macrocycle generation, but they are not without drawbacks. Our current work has focused on the CREMP dataset, which is limited to homodetic, 4-, 5-, and 6-mer macrocycles with canonical side chains. Extension to macrocycles with larger ring sizes, non-canonical side chains, and other complex topologies would improve the generalizability of this work, as well as training on datasets generated with higher levels of theory. Additionally, although we demonstrate the effectiveness of a standard, discrete-time diffusion process, our approach is not physically constrained to satisfy macrocyclic geometries and currently requires a post-optimization step. Developing and applying physics-informed diffusion processes with manifold constraints could improve the efficiency of training and sampling of relevant macrocycle backbones.

## 6 CONCLUSIONS

We present RINGER, a new approach for generating macrocycle conformer ensembles that significantly improves sample quality, diversity, and inference. By leveraging strengths of diffusion-based models, we demonstrate how a transformer-based architecture with a cyclic positional encoding results in significant gains over Cartesian-based equivariant models and widely-used distance geometry-based algorithms for both unconditional and conditional structure generation. The present work paves the way for more efficient and accurate computational exploration of conformational space. We anticipate that this approach will more broadly enable rational macrocycle discovery through further development.

## REFERENCES

- Hiba Alogheli, Gustav Olanders, Wesley Schaal, Peter Brandt, and Anders Karlén. Docking of macrocycles: Comparing rigid and flexible docking in glide. *J. Chem. Inf. Model.*, 57(2):190–202, February 2017.
- Namrata Anand and Tudor Achim. Protein structure and sequence generation with equivariant denoising diffusion probabilistic models, 2022. URL <https://arxiv.org/abs/2205.15019>.
- Simon Axelrod and Rafael Gómez-Bombarelli. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Sci. Data*, 9(1):185, April 2022.
- Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N Kinch, R Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R Glassman, Andy DeGiovanni, Jose H Pereira, Andria V Rodrigues, Alberdina A van Dijk, Ana C Ebrecht, Diederik J Opperman, Theo Sagmeister, Christoph Buhheller, Tea Pavkov-Keller, Manoj K Rathinaswamy, Udit Dalwadi, Calvin K Yip, John E Burke, K Christopher Garcia, Nick V Grishin, Paul D Adams, Randy J Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, August 2021.
- Christoph Bannwarth, Sebastian Ehlert, and Stefan Grimme. GFN2-xTB—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671, March 2019.
- H M Berman, J Westbrook, Z Feng, G Gilliland, T N Bhat, H Weissig, I N Shindyalov, and P E Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–242, January 2000.
- 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, September 2022.
- George Chang, Wayne C Guida, and W Clark Still. An internal-coordinate monte carlo method for searching conformational space. *J. Am. Chem. Soc.*, 111(12):4379–4386, June 1989.
- David J Craik, David P Fairlie, Spiros Liras, and David Price. The future of peptide-based drugs. *Chem. Biol. Drug Des.*, 81(1):136–147, January 2013.
- Jovan Damjanovic, Jiayuan Miao, He Huang, and Yu-Shan Lin. Elucidating solution structures of cyclic peptides using molecular dynamics simulations. *Chem. Rev.*, 121(4):2292–2324, February 2021.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. URL <https://aclanthology.org/N19-1423>.
- Edward M Driggers, Stephen P Hale, Jinbo Lee, and Nicholas K Terrett. The exploration of macrocycles for drug discovery—an underexploited structural class. *Nat. Rev. Drug Discov.*, 7(7):608–624, July 2008.
- Roland L. Dunbrack and Martin Karplus. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat. Struct. Biol.*, 1(5):334–340, 1994.
- William Falcon and The PyTorch Lightning team. PyTorch Lightning, March 2019. URL <https://github.com/Lightning-AI/lightning>.

- Nic Fishman, Leo Klarner, Valentin De Bortoli, Emile Mathieu, and Michael Hutchinson. Diffusion models for constrained domains, 2023. URL <https://arxiv.org/abs/2304.05364>.
- Octavian-Eugen Ganea, Lagnajit Pattanaik, Connor W. Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi S. Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. In *Advances in Neural Information Processing Systems*, 2021. URL [https://openreview.net/forum?id=af\\_hng9tuNj](https://openreview.net/forum?id=af_hng9tuNj).
- Diego Garcia Jimenez, Vasanthanathan Poongavanam, and Jan Kihlberg. Macrocycles in drug discovery - learning from the past for the future. *J. Med. Chem.*, 66:5377–5396, April 2023.
- Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=BlEwBxStPH>.
- Ross Girshick. Fast R-CNN. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015. doi: 10.1109/ICCV.2015.169.
- Colin A Grambow, Hayley Weir, Christian N Cunningham, Tommaso Biancalani, and Kangway V Chuang. CREMP: Conformer-rotamer ensembles of macrocyclic peptides for machine learning, 2023. URL <https://arxiv.org/abs/2305.08057>.
- Colin R Groom, Ian J Bruno, Matthew P Lightfoot, and Suzanna C Ward. The cambridge structural database. *Acta. Crystallogr. B. Struct. Sci. Cryst. Eng. Mater.*, 72(Pt 2):171–179, April 2016.
- Thomas A Halgren. Merck molecular force field. V. Extension of MMFF94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.*, 17(5–6):616–641, 1996.
- Li Han and Lee Rudolph. Inverse kinematics for a serial chain with joints under distance constraints. In *Robotics: Science and Systems*, 2006.
- Paul C D Hawkins and Anthony Nicholls. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J. Chem. Inf. Model.*, 52(11):2919–2936, November 2012.
- 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. *J. Chem. Inf. Model.*, 50(4):572–584, April 2010.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020. URL [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf).
- 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. Curran Associates, Inc., 2022. URL [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/994545b2308bbbbc97e3e687ea9e464f-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/994545b2308bbbbc97e3e687ea9e464f-Paper-Conference.pdf).
- 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, August 2021.
- Anna S Kamenik, Uta Lessel, Julian E Fuchs, Thomas Fox, and Klaus R Liedl. Peptidic macrocycles - conformational sampling and thermodynamic characterization. *J. Chem. Inf. Model.*, 58(5): 982–992, May 2018.

- István Kolossváry and Wayne C Guida. Low mode search. an efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *J. Am. Chem. Soc.*, 118(21):5011–5019, January 1996.
- István Kolossváry and Wayne C Guida. Low-mode conformational search elucidated: Application to C39H80 and flexible docking of 9-deazaguanine inhibitors into PNP. *J. Comput. Chem.*, 20(15):1671–1684, November 1999.
- Dieter Kraft. A software package for sequential quadratic programmings. *Tech. Rep. DFVLR-FB 88-28*, 1988.
- Greg Landrum. RDKit: Open-source cheminformatics, 2006. URL <https://rdkit.org>.
- Siegfried S F Leung, Daniel Sindhikara, and Matthew P Jacobson. Simple predictive models of passive membrane permeability incorporating size-dependent membrane-water partition. *J. Chem. Inf. Model.*, 56(5):924–929, May 2016.
- Stephanie M Linker, Christian Schellhaas, Anna S Kamenik, Mac M Veldhuizen, Franz Waibl, Hans-Jörg Roth, Marianne Fouché, Stephane Rodde, and Sereina Riniker. Lessons for oral bioavailability: How conformationally flexible cyclic peptides enter and cross lipid membranes. *J. Med. Chem.*, 66(4):2773–2788, February 2023.
- Elman Mansimov, Omar Mahmood, Seokho Kang, and Kyunghyun Cho. Molecular geometry prediction using a deep generative graph neural network. *Sci. Rep.*, 9(1):20381, December 2019.
- Jiayuan Miao, Marc L Descoteaux, and Yu-Shan Lin. Structure prediction of cyclic peptides by molecular dynamics + machine learning. *Chem. Sci.*, 12(44):14927–14936, November 2021.
- Markus Muttenthaler, Glenn F King, David J Adams, and Paul F Alewood. Trends in peptide drug discovery. *Nat. Rev. Drug Discov.*, 20(4):309–325, April 2021.
- Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8162–8171. PMLR, 18–24 Jul 2021. URL <https://proceedings.mlr.press/v139/nichol21a.html>.
- Daniel S Nielsen, Nicholas E Shepherd, Weijun Xu, Andrew J Lucke, Martin J Stoermer, and David P Fairlie. Orally absorbed cyclic peptides. *Chem. Rev.*, 117(12):8094–8128, June 2017.
- OpenEye. OMEGA 4.2. 1.2, 2022. URL <http://www.eyesopen.com>.
- Jerod Parsons, J Bradley Holmes, J Maurice Rojas, Jerry Tsai, and Charlie E M Strauss. Practical conversion from torsion space to cartesian space for in silico protein synthesis. *J. Comput. Chem.*, 26(10):1063–1068, 2005.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.
- Chunyang Peng, Philippe Y. Ayala, H. Bernhard Schlegel, and Michael J. Frisch. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. of Comput. Chem.*, 17(1):49–56, 1996.
- Vasanthanathan Poongavanam, Emma Danelius, Stefan Peintner, Lilian Alcaraz, Giulia Caron, Maxwell D Cummings, Stanislaw Wlodek, Mate Erdelyi, Paul C D Hawkins, Giuseppe Ermondi, and Jan Kihlberg. Conformational sampling of macrocyclic drugs in different environments: Can we find the relevant conformations? *ACS Omega*, 3(9):11742–11757, September 2018.
- Philipp Pracht, Fabian Bohle, and Stefan Grimme. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.*, 22(14):7169–7192, April 2020.

- G N Ramachandran and V Sasisekharan. Conformation of polypeptides and proteins. *Adv. Protein Chem.*, 23:283–438, 1968.
- Taha Rezai, Jonathan E Bock, Mai V Zhou, Chakrapani Kalyanaraman, R Scott Lokey, and Matthew P Jacobson. Conformational flexibility, internal hydrogen bonding, and passive membrane permeability: successful in silico prediction of the relative permeabilities of cyclic peptides. *J. Am. Chem. Soc.*, 128(43):14073–14080, November 2006.
- Sereina Riniker and Gregory A Landrum. Better informed distance geometry: Using what we know to improve conformation generation. *J. Chem. Inf. Model.*, 55(12):2562–2574, December 2015.
- Christin Schärfer, Tanja Schulz-Gasch, Jérôme Hert, Lennart Heinzerling, Benjamin Schulz, Therese Inhester, Martin Stahl, and Matthias Rarey. CONFECT: conformations from an expert collection of torsion patterns. *ChemMedChem*, 8(10):1690–1700, October 2013.
- Tanja Schulz-Gasch, Christin Schärfer, Wolfgang Guba, and Matthias Rarey. TFD: Torsion fingerprints as a new measure to compare small molecule conformations. *J. Chem. Inf. Model.*, 52(6):1499–1512, June 2012.
- Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 464–468, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. URL <https://aclanthology.org/N18-2074>.
- Gregor Simm and Jose Miguel Hernandez-Lobato. A generative model for molecular distance geometry. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8949–8958. PMLR, 13–18 Jul 2020. URL <https://proceedings.mlr.press/v119/simm20a.html>.
- 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. *J. Chem. Inf. Model.*, 57(8):1881–1894, August 2017.
- Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=PxTIG12RRHS>.
- D C Spellmeyer, A K Wong, M J Bower, and J M Blaney. Conformational analysis using distance geometry methods. *J. Mol. Graph. Model.*, 15(1):18–36, February 1997.
- Hannes Stärk, Octavian Ganea, Lagnajit Pattanaik, Regina Barzilay, and Tommi Jaakkola. EquiBind: Geometric deep learning for drug binding structure prediction. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 20503–20521. PMLR, 17–23 Jul 2022. URL <https://proceedings.mlr.press/v162/stark22b.html>.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- Elizabeth A Villar, Dmitri Beglov, Spandan Chennamadhavuni, John A Porco, Jr, Dima Kozakov, Sandor Vajda, and Adrian Whitty. How proteins bind macrocycles. *Nat. Chem. Biol.*, 10(9):723–731, September 2014.
- Alexander A Vinogradov, Yizhen Yin, and Hiroaki Suga. Macrocyclic peptides as drug candidates: Recent progress and remaining challenges. *J. Am. Chem. Soc.*, 141(10):4167–4181, March 2019.

- 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 preferences. *J. Chem. Inf. Model.*, 60(4):2044–2058, April 2020.
- Shuzhe Wang, Kajo Krummenacher, Gregory A Landrum, Benjamin D Sellers, Paola Di Lello, Sarah J Robinson, Bryan Martin, Jeffrey K Holden, Jeffrey Y K Tom, Anastasia C Murthy, Nataliya Popovych, and Sereina Riniker. Incorporating NOE-Derived distances in conformer generation of cyclic peptides with distance geometry. *J. Chem. Inf. Model.*, 62(3):472–485, February 2022.
- K Shawn Watts, Pranav Dalal, Andrew J Tebben, Daniel L Cheney, and John C Shelley. Macrocycle conformational sampling with MacroModel. *J. Chem. Inf. Model.*, 54(10):2680–2696, October 2014.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Online, October 2020. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/2020.emnlp-demos.6>.
- Kevin E. Wu, Kevin K. Yang, Rianne van den Berg, James Y. Zou, Alex X. Lu, and Ava P. Amini. Protein structure generation via folding diffusion, 2022. URL <https://arxiv.org/abs/2209.15611>.
- Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=pAbmlqfheGk>.
- 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://openreview.net/forum?id=PzcvxEMzvQC>.
- Jason Yim, Brian L. Trippe, Valentin De Bortoli, Emile Mathieu, Arnaud Doucet, Regina Barzilay, and Tommi Jaakkola. SE(3) diffusion model with application to protein backbone generation, 2023. URL <https://arxiv.org/abs/2302.02277>.
- 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 M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 28877–28888. Curran Associates, Inc., 2021. URL [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/f1c1592588411002af340cbaedd6fc33-Paper.pdf).
- 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. ISSN 2835-8856. URL <https://openreview.net/forum?id=lCPOHiztuw>.