
RINGER: Rapid Conformer Generation for Macrocycles with Sequence-Conditioned Internal Coordinate Diffusion

Anonymous Authors¹

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 for sequence-conditioned generation of macrocycle structures based on internal coordinates. RINGER provides fast backbone 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 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 (Craik 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 sig-

nificantly 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) Macrocylic peptides exhibit diverse molecular structures and chemical modifications which results in a vast conformational space (Kamenik et al., 2018). 2) Macrocycles are subject to complex non-linear constraints due to ring closure (Watts et al., 2014). 3) Experimental structures for macrocycles are lacking ($\sim 10^3$), making it difficult to improve structural predictions or train machine learning-based approaches. To address these limitations, we introduce RINGER (RINGER Generates Ensembles of Rings), a deep learning model designed for sequence-conditioned macrocycle conformer generation (Figure 1) that efficiently samples realistic angles and torsions (i.e., internal coordinates). RINGER merges a transformer architecture with a discrete-time diffusion model. We demonstrate the strong performance of RINGER in sample quality, diversity and RMSDs relative to gold-standard conformer ensembles generated with CREST (Pracht et al., 2020). The contributions of this work are summarized as follows: (i) We develop a new framework for conformer generation of macrocycle backbones based on efficiently encoding ring geometry using redundant internal coordinates. (ii) We propose a simple solution to recover Cartesian coordinates from redundant internal coordinates that satisfies ring constraints. (iii) We benchmark RINGER extensively against state-of-the-art physics- and machine learning-based algorithms to demonstrate the excellent model performance.

2. Background and Related Work

Physics and Heuristic-based Conformer Generation

Physics-based and heuristic-based algorithms remain the state of the art for macrocycles and have required special considerations due to ring-closing constraints. RDKit leverages distance geometry algorithms for macrocycle conformer generation (ETKDGv3) (Riniker & Landrum, 2015; Wang et al., 2020; 2022). Similarly, commercial conformer generation algorithms such as OpenEye OMEGA (Hawkins

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

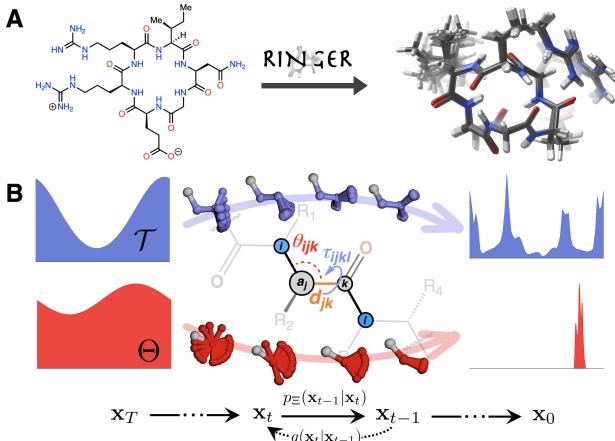


Figure 1. Overview of RINGER for macrocycle conformation generation. **A.** Given a 2D representation of a macrocyclic peptide, RINGER generates an accurate and diverse 3D conformational ensemble. **B.** An illustration of the diffusion process learning to recover the time $t = 0$ bond angle (red) and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

et al., 2010; Hawkins & Nicholls, 2012) in macrocycle mode use a distance geometry algorithm based on 4D coordinate initialization (Spellmeyer et al., 1997). The open-source CREST package (Pracht et al., 2020) leverages iterative metadynamics with a genetic structure-crossing algorithm (iMTD-GC), and can be considered a gold-standard for generating diverse ensembles.

Machine Learning for Conformer Generation Recent work with deep generative models has focused on improved sampling of the conformational landscape of small molecules (Mansimov et al., 2019; Simm & Hernandez-Lobato, 2020; Xu et al., 2021). GeoDiff, reported by Xu et al. (2022), is 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. Other approaches have drawn inspiration from physics-based conformer generation (Schäfer et al., 2013; Ganea et al., 2021; Stärk et al., 2022). Recently, Jing et al. (2022) report Torsional Diffusion, a diffusion model that operates on the torsional space via an extrinsic-to-intrinsic score model. Importantly, these methods do not address the challenge of highly coupled torsions within cyclic systems.

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). Our work builds on FoldingDiff, parameterizing structures over internal backbone angles and torsions. However, we account for the cyclic constraints of macrocycles and focus instead on conditional 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. However, 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 α -carbon representation.

3. RINGER: Problem Statement and Methods

3.1. Problem Definition: Conditional Conformer Generation

Given a macrocycle graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{V} is the set of n nodes (atoms) and \mathcal{E} is the set of edges (bonds), our task is to learn $p(\mathcal{C} | \mathcal{G})$, 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. Given this problem, a good generative model should naturally encode the cyclic shift invariance of macrocycles and capture multimodal distributions and complex, higher-order interactions to produce high-quality and diverse ensembles.

3.2. Representing Macrocycle Geometry

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). 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 ring chirality. Moreover, this approach obviates the need for complex equivariant networks and enables the use of simpler neural architectures (Wu et al., 2022). Hence, our

generative process can be reformulated as learning the distribution $p(\{\Theta, \mathcal{T}\} | \mathcal{G}; \mathcal{D})$ using known bond distances for reconstruction back to Cartesians (Figure 1).

3.3. Diffusion Models for Internal Coordinates

Our model represents macrocycle backbones as cyclic sequences of redundant angles and dihedrals with fixed bond lengths. We train a discrete-time diffusion model (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021) from Wu et al. (2022) that formulates the forward transition probability using a wrapped normal distribution to predict the noise at a given time step. (Jing et al., 2022). For full details, see Appendix C. At inference time, we sample from a wrapped Gaussian distribution to produce a set of angles and torsions, conditioning on the known set of atom features corresponding to the amino-acid sequence. The sampling process is further detailed in Appendix D. We apply a bidirectional transformer architecture (Vaswani et al., 2017; Devlin et al., 2019) with an invariant cyclic positional encoding (Shaw et al., 2018) (see Appendix A for details). For conditional generation, we perform a linear projection of the features \mathbf{a}_i , corresponding to each macrocycle backbone atom and its side chain, and a separate linear projection of the angles and torsions $\mathbf{x}_i = [\theta_i, \tau_i]$ 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 . Model details are shown in Appendix E.

Macrocycles with fixed bond distances contain three redundant torsional angles and two redundant bond angles. These redundancies prevent direct transformation to unique Cartesians for cyclic structures. Adopting a sequential reconstruction method such as NeRF (Parsons et al., 2005) accumulates small errors that result in inadequate ring closure for macrocycles. 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. Here, we find the set of Cartesian coordinates that minimize the squared error against the internal coordinates sampled by the diffusion process while satisfying bond distance equality constraints using known bond distances from the training data. 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).

4. Experiments and Results

4.1. Dataset, Training & Sampling

We train and evaluate our approach on the recently published CREMP dataset (Grambow et al., 2023) that contains 36k homodetic macrocyclic peptides, using 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). Full dataset statistics are shown in Appendix B. We train each model for up to 1000 epochs until convergence (typically less than 100 epochs) 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.

4.2. Unconditional Generation of Macrocycles

To understand whether this approach can learn the underlying distribution of macrocycle conformations, we first trained RINGER on macrocycle backbones in the absence of any atom or side-chain features and only providing ring-atom identity. From a design perspective, diverse backbone sampling alone can help drive inverse peptide design, where specific backbone geometries suggest important sequences. Figure 2 clearly demonstrates how RINGER accurately replicates both angles and dihedrals with tight fidelity across all residue atoms, both qualitatively from the plots and quantitatively as measured by the KL divergence. 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. Appendix K provides more fine-grained detail by visualizing distributions separately based on the number of residues in the macrocycle.

4.3. Sequence-Conditioned Generation of Macrocycles

Whereas our unconditional model above disregarded side chains, we now condition backbone generation on molecular features corresponding to each residue, including side-chain features, stereochemistry, and *N*-methylation (see Appendix E). Comparison of RINGER RMSD and TFD ensemble metrics (see Appendix I for definitions) against RDKit, OMEGA, and GeoDiff 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 (also see Appendix L.4 for confidence intervals). We found that RDKit ETKDGv3 and OMEGA Macrocyclic mode performed similarly across both metrics and achieved moderate recall with limited precision. To compare deep learning approaches, we trained a Euclidean diffusion model using the GeoDiff architecture (GeoDiff-Macro) on the CREMP dataset, and found a strong boost in recall with similar precision.

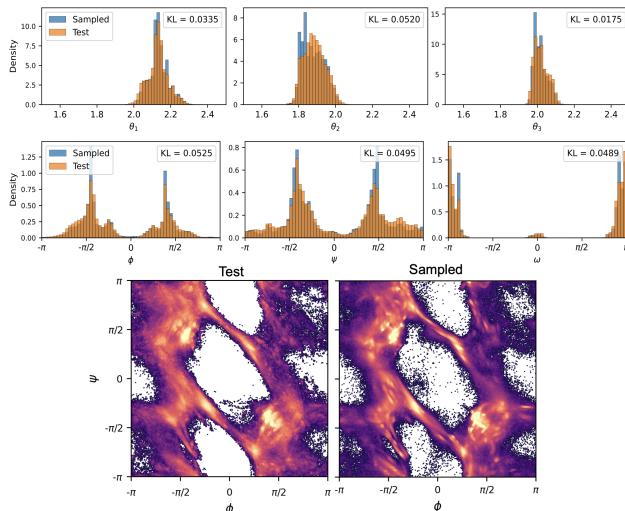


Figure 2. Comparison of the bond angle and dihedral distributions from the held-out test set (orange) and in the unconditionally generated samples (blue). Showing the three bond angle types in each amino acid residue (top), the three dihedral angles for each residue (middle), and the Ramachandran plots (bottom, colored logarithmically by density with high density regions shown in lighter colors). KL divergence is calculated as $D_{KL}(\text{test} \parallel \text{sampled})$.

We evaluated our approach with and without the post-processing geometry constrained optimization (Equation 6), as our raw, generated samples from RINGER may not satisfy macrocycle distance constraints. As with unconditional generation, sequence-conditioned generation learns the data distribution with high fidelity as shown in Appendix L.1. “RINGER” in Table 1 corresponds to Cartesian geometries that were generated from the predicted angles by starting at one atom and setting bond distances, angles, and dihedrals sequentially. “RINGER (opt)” refers to our post-processed geometries that satisfy the true bond distances exactly. Notably, both approaches achieve excellent recall and precision across both RMSD- and TFD-based scores compared to our baselines. Furthermore, post-processing to guarantee valid macrocycle geometries preserves excellent recall, albeit with slightly attenuated precision. Importantly, our approach provides increased sampling efficiency with only $T = 20$ time steps, compared to GeoDiff’s $T = 5000$ or FoldingDiff’s $T = 1000$ (see Table 7 in Appendix L.5).

4.4. Structural Analysis of Generated Macrocycles

Although RMSD and TFD give a quantitative evaluation of performance, we also analyzed individual ensembles to understand the qualitative differences in conformer generation processes (Figure 3). Notably, the two macrocycles shown possess distinct sequences that result in distinct Ramachandran plots and conformations. As shown in Figure 3 (top panels), most ground truth conformer ensembles exhibit relatively tight distributions characterized by a distinctive set

Table 1. Mean performance metrics for sequence-conditioned generation of macrocycles evaluated on ring atoms. Coverage is evaluated at a threshold of 0.1 Å for RMSD and 0.05 for TFD. Coverage is reported as % values and RMSD-MAT is reported in Å. All test data conformers are used for evaluation. “opt” refers to the use of Equation (6) for reconstruction.

METHOD	RMSD		PRECISION	
	RECALL COV ↑	PRECISION MAT ↓	PRECISION Cov. ↑	PRECISION MAT ↓
RDKIT	35.8	0.187	5.6	0.540
OMEGA	32.3	0.186	3.7	0.557
GEOIFF-MACRO	50.8	0.151	6.4	0.592
RINGER	77.0	0.091	61.3	0.185
RINGER (OPT)	79.7	0.084	56.4	0.441
TFD				
RDKIT	52.9	0.059	9.4	0.215
OMEGA	49.7	0.061	6.6	0.225
GEOIFF-MACRO	68.1	0.048	9.1	0.248
RINGER	90.1	0.024	74.7	0.059
RINGER (OPT)	89.2	0.024	61.8	0.068

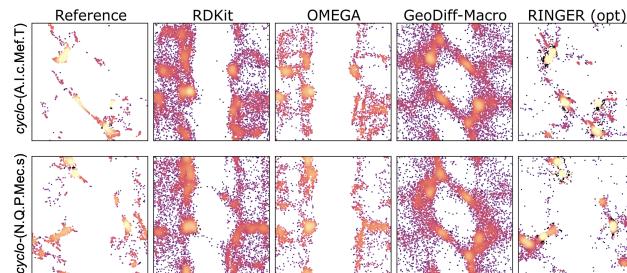


Figure 3. RINGER accurately generates ensembles as illustrated by Ramachandran plots for individual macrocycle ensembles.

of ϕ, ψ angles. Although RDKit, OMEGA, and GeoDiff can identify relevant low-energy conformers (albeit with slight errors, see Appendix L.2 for examples), the overall sampling process generates unrealistic distributions. In contrast, RINGER recapitulates not only the ground state geometry with excellent accuracy (Figure 11), but better captures the entire ensemble distribution.

5. Conclusions and Future Directions

In summary, we present RINGER, a new approach for generating macrocycle conformer ensembles that significantly improves sample quality, diversity, and inference. Our studies demonstrate how diffusion-based models can tackle limitations in constrained generation and are an important step toward efficient generation of complex macrocycle geometries. Ongoing work is focused on 1) extending our approach to macrocycles with larger ring sizes, non-canonical side chains and fully modeling the positions of all side chain atom; and 2) developing physics-informed diffusion processes that inherently incorporate inequality constraints.

References

- 220 Alogheli, H., Olanders, G., Schaal, W., Brandt, P., and
 221 Karlén, A. Docking of macrocycles: Comparing rigid
 222 and flexible docking in glide. *J. Chem. Inf. Model.*, 57(2):
 223 190–202, February 2017.
- 224 Anand, N. and Achim, T. Protein structure and sequence
 225 generation with equivariant denoising diffusion proba-
 226 bility models, 2022. URL <https://arxiv.org/abs/2205.15019>.
- 227 Bhardwaj, G., O'Connor, J., Rettie, S., Huang, Y.-H.,
 228 Ramelot, T. A., Mulligan, V. K., Alpkilic, G. G., Palmer,
 229 J., Bera, A. K., Bick, M. J., Di Piazza, M., Li, X., Hos-
 230 seinzadeh, P., Craven, T. W., Tejero, R., Lauko, A., Choi,
 231 R., Glynn, C., Dong, L., Griffin, R., van Voorhis, W. C.,
 232 Rodriguez, J., Stewart, L., Montelione, G. T., Craik, D.,
 233 and Baker, D. Accurate de novo design of membrane-
 234 traversing macrocycles. *Cell*, 185(19):3520–3532.e26,
 235 September 2022.
- 236 Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. The future
 237 of peptide-based drugs. *Chem. Biol. Drug Des.*, 81(1):
 238 136–147, January 2013.
- 239 Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. BERT:
 240 Pre-training of deep bidirectional transformers for lan-
 241 guage understanding. In *Proceedings of the 2019 Con-
 242 ference of the North American Chapter of the Associa-
 243 tion for Computational Linguistics: Human Language
 244 Technologies, Volume 1 (Long and Short Papers)*, pp.
 245 4171–4186, Minneapolis, Minnesota, June 2019. Asso-
 246 ciation for Computational Linguistics. URL <https://aclanthology.org/N19-1423>.
- 247 Driggers, E. M., Hale, S. P., Lee, J., and Terrett, N. K.
 248 The exploration of macrocycles for drug discovery—an
 249 underexploited structural class. *Nat. Rev. Drug Discov.*, 7
 250 (7):608–624, July 2008.
- 251 Falcon, W. and The PyTorch Lightning team. PyTorch Light-
 252 ning, March 2019. URL <https://github.com/Lightning-AI/lightning>.
- 253 Fishman, N., Klarner, L., Bortoli, V. D., Mathieu, E., and
 254 Hutchinson, M. Diffusion models for constrained
 255 domains, 2023. URL <https://arxiv.org/abs/2304.05364>.
- 256 Ganea, O.-E., Pattanaik, L., Coley, C. W., Barzilay, R.,
 257 Jensen, K., Green, W., and Jaakkola, T. S. Geomol: Tor-
 258 sional geometric generation of molecular 3d conformer
 259 ensembles. In *Advances in Neural Information Pro-
 260 cessing Systems*, 2021. URL https://openreview.net/forum?id=af_hng9tuNj.
- 261 Garcia Jimenez, D., Poongavanam, V., and Kihlberg, J.
 262 Macrocycles in drug discovery - learning from the past
 263 for the future. *J. Med. Chem.*, 66:5377–5396, April 2023.
- 264 Gasteiger, J., Groß, J., and Günemann, S. Directional
 265 message passing for molecular graphs. In *Inter-
 266 national Conference on Learning Representations*, 2020. URL <https://openreview.net/forum?id=BleWbxStPH>.
- 267 Girshick, R. Fast R-CNN. In *2015 IEEE International
 268 Conference on Computer Vision (ICCV)*, pp. 1440–1448,
 269 2015. doi: 10.1109/ICCV.2015.169.
- 270 Grambow, C. A., Weir, H., Cunningham, C. N., Biancalani,
 271 T., and Chuang, K. V. CREMP: Conformer-rotamer
 272 ensembles of macrocyclic peptides for machine learn-
 273 ing, 2023. URL <https://arxiv.org/abs/2305.08057>.
- 274 Halgren, T. A. Merck molecular force field. V. Extension
 275 of MMFF94 using experimental data, additional compu-
 276 tational data, and empirical rules. *J. Comput. Chem.*, 17
 277 (5–6):616–641, 1996.
- 278 Hawkins, P. C. D. and Nicholls, A. Conformer generation
 279 with OMEGA: learning from the data set and the analysis
 280 of failures. *J. Chem. Inf. Model.*, 52(11):2919–2936,
 281 November 2012.
- 282 Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Elling-
 283 son, B. A., and Stahl, M. T. Conformer generation with
 284 OMEGA: algorithm and validation using high quality
 285 structures from the protein databank and cambridge struc-
 286 tural database. *J. Chem. Inf. Model.*, 50(4):572–584, April
 287 2010.
- 288 Ho, J., Jain, A., and Abbeel, P. Denoising diffu-
 289 sion probabilistic models. In *Advances in Neu-
 290 ral Information Processing Systems*, volume 33,
 291 pp. 6840–6851. Curran Associates, Inc., 2020.
 292 URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.
- 293 Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T.
 294 Torsional diffusion for molecular conformer generation.
 295 In *Advances in Neural Information Processing Systems*,
 296 volume 35, pp. 24240–24253. Curran Associates, Inc.,
 297 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/994545b2308bbbbbc97e3e687ea9e464f-Paper-Conference.pdf.

- 275 Kamenik, A. S., Lessel, U., Fuchs, J. E., Fox, T., and
 276 Liedl, K. R. Peptidic macrocycles - conformational sam-
 277 pling and thermodynamic characterization. *J. Chem. Inf.*
 278 *Model.*, 58(5):982–992, May 2018.
- 279 Kraft, D. A software package for sequential quadratic pro-
 280 grammings. *Tech. Rep. DFVLR-FB* 88-28, 1988.
- 281 Landrum, G. RDKit: Open-source cheminformatics, 2006.
 282 URL <https://rdkit.org>.
- 283 Leung, S. S. F., Sindhikara, D., and Jacobson, M. P. Simple
 284 predictive models of passive membrane permeability incor-
 285 porating size-dependent membrane-water partition. *J.*
 286 *Chem. Inf. Model.*, 56(5):924–929, May 2016.
- 287 Mansimov, E., Mahmood, O., Kang, S., and Cho, K. Molecular
 288 geometry prediction using a deep generative graph
 289 neural network. *Sci. Rep.*, 9(1):20381, December 2019.
- 290 Miao, J., Descoteaux, M. L., and Lin, Y.-S. Structure predic-
 291 tion of cyclic peptides by molecular dynamics + machine
 292 learning. *Chem. Sci.*, 12(44):14927–14936, November
 293 2021.
- 294 Mutenthaler, M., King, G. F., Adams, D. J., and Alewood,
 295 P. F. Trends in peptide drug discovery. *Nat. Rev. Drug*
 296 *Discov.*, 20(4):309–325, April 2021.
- 297 Nichol, A. Q. and Dhariwal, P. Improved denoising
 298 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
 299 2021. URL <https://proceedings.mlr.press/v139/nichol21a.html>.
- 300 Nielsen, D. S., Shepherd, N. E., Xu, W., Lucke, A. J., Sto-
 301 ermer, M. J., and Fairlie, D. P. Orally absorbed cyclic
 302 peptides. *Chem. Rev.*, 117(12):8094–8128, June 2017.
- 303 Parsons, J., Bradley Holmes, J., Maurice Rojas, J., Tsai, J.,
 304 and Strauss, C. E. M. Practical conversion from torsion
 305 space to cartesian space for in silico protein synthesis. *J.*
 306 *Comput. Chem.*, 26(10):1063–1068, 2005.
- 307 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J.,
 308 Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga,
 309 L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison,
 310 M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L.,
 311 Bai, J., and Chintala, S. Pytorch: An imperative style,
 312 high-performance deep learning library. In *Advances in Neural Information Processing Systems* 32, pp. 8024–
 313 8035. Curran Associates, Inc., 2019.
- 314 Poongavanam, V., Danelius, E., Peintner, S., Alcaraz, L.,
 315 Caron, G., Cummings, M. D., Wlodek, S., Erdelyi, M.,
 316 Hawkins, P. C. D., Ermondi, G., and Kihlberg, J. Con-
 317 formational sampling of macrocyclic drugs in different
 318 environments: Can we find the relevant conformations?
 319 *ACS Omega*, 3(9):11742–11757, September 2018.
- 320 Pracht, P., Bohle, F., and Grimme, S. Automated explo-
 321 ration of the low-energy chemical space with fast quan-
 322 tum chemical methods. *Phys. Chem. Chem. Phys.*, 22
 323 (14):7169–7192, April 2020.
- 324 Ramachandran, G. N. and Sasisekharan, V. Conformation
 325 of polypeptides and proteins. *Adv. Protein Chem.*, 23:
 326 283–438, 1968.
- 327 Rezai, T., Bock, J. E., Zhou, M. V., Kalyanaraman, C.,
 328 Lokey, R. S., and Jacobson, M. P. Conformational flexi-
 329 bility, internal hydrogen bonding, and passive membrane
 330 permeability: successful in silico prediction of the rela-
 331 tive permeabilities of cyclic peptides. *J. Am. Chem. Soc.*,
 332 128(43):14073–14080, November 2006.
- 333 Riniker, S. and Landrum, G. A. Better informed distance
 334 geometry: Using what we know to improve conformation
 335 generation. *J. Chem. Inf. Model.*, 55(12):2562–2574,
 336 December 2015.
- 337 Schärfer, C., Schulz-Gasch, T., Hert, J., Heinzerling, L.,
 338 Schulz, B., Inhester, T., Stahl, M., and Rarey, M. CON-
 339 FECT: conformations from an expert collection of torsion
 340 patterns. *ChemMedChem*, 8(10):1690–1700, October
 341 2013.
- 342 Schulz-Gasch, T., Schärfer, C., Guba, W., and Rarey, M.
 343 TFD: Torsion fingerprints as a new measure to compare
 344 small molecule conformations. *J. Chem. Inf. Model.*, 52
 345 (6):1499–1512, June 2012.
- 346 Shaw, P., Uszkoreit, J., and Vaswani, A. Self-attention
 347 with relative position representations. In *Proceedings
 348 of the 2018 Conference of the North American Chapter
 349 of the Association for Computational Linguistics: Hu-
 350 man Language Technologies, Volume 2 (Short Papers)*,
 351 pp. 464–468, New Orleans, Louisiana, June 2018. As-
 352 sociation for Computational Linguistics. URL <https://aclanthology.org/N18-2074>.
- 353 Simm, G. and Hernandez-Lobato, J. M. A genera-
 354 tive model for molecular distance geometry. In *Pro-
 355 ceedings of the 37th International Conference on Ma-
 356 chine Learning*, volume 119 of *Proceedings of Machine
 357 Learning Research*, pp. 8949–8958. PMLR, 13–18 Jul
 358 2020. URL <https://proceedings.mlr.press/v119/simm20a.html>.
- 359 Software, O. S. OMEGA 4.2. 1.2, 2022. URL <http://www.eyesopen.com>.

- 330 Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and
331 Ganguli, S. Deep unsupervised learning using nonequilibrium
332 thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*,
333 volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.
334 URL <https://proceedings.mlr.press/v37/sohl-dickstein15.html>.
- 335 Watts, K. S., Dalal, P., Tebben, A. J., Cheney, D. L., and
336 Shelley, J. C. Macrocyclic conformational sampling with
337 MacroModel. *J. Chem. Inf. Model.*, 54(10):2680–2696,
338 October 2014.
- 339 Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue,
340 Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz,
341 M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite,
342 Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame,
343 M., Lhoest, Q., and Rush, A. M. Transformers: State-
344 of-the-art natural language processing. In *Proceedings
345 of the 2020 Conference on Empirical Methods in Natural
346 Language Processing: System Demonstrations*, pp.
347 38–45, Online, October 2020. Association for Computational
348 Linguistics. URL <https://www.aclweb.org/anthology/2020.emnlp-demos.6>.
- 349 Spellmeyer, D. C., Wong, A. K., Bower, M. J., and Blaney,
350 J. M. Conformational analysis using distance geometry
351 methods. *J. Mol. Graph. Model.*, 15(1):18–36, February
352 1997.
- 353 Stärk, H., Ganea, O., Pattanaik, L., Barzilay, R., and
354 Jaakkola, T. EquiBind: Geometric deep learning for
355 drug binding structure prediction. In *Proceedings of the 39th International Conference on Machine Learning*,
356 volume 162 of *Proceedings of Machine Learning Research*, pp. 20503–20521. PMLR, 17–23 Jul
357 2022. URL <https://proceedings.mlr.press/v162/stark22b.html>.
- 358 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones,
359 L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. Attention
360 is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.,
361 2017.
- 362 Villar, E. A., Beglov, D., Chennamadhavuni, S., Porco,
363 Jr, J. A., Kozakov, D., Vajda, S., and Whitty, A. How
364 proteins bind macrocycles. *Nat. Chem. Biol.*, 10(9):723–
365 731, September 2014.
- 366 Vinogradov, A. A., Yin, Y., and Suga, H. Macrocyclic
367 peptides as drug candidates: Recent progress and remain-
368 ing challenges. *J. Am. Chem. Soc.*, 141(10):4167–4181,
369 March 2019.
- 370 Wang, S., Witek, J., Landrum, G. A., and Riniker, S. Improv-
371 ing conformer generation for small rings and macrocycles
372 based on distance geometry and experimental torsional-
373 angle preferences. *J. Chem. Inf. Model.*, 60(4):2044–
374 2058, April 2020.
- 375 Wang, S., Krummenacher, K., Landrum, G. A., Sellers,
376 B. D., Di Lello, P., Robinson, S. J., Martin, B., Holden,
377 J. K., Tom, J. Y. K., Murthy, A. C., Popovych, N., and
378 Riniker, S. Incorporating NOE-Derived distances in con-
379 former generation of cyclic peptides with distance geo-
380 metry. *J. Chem. Inf. Model.*, 62(3):472–485, February
381 2022.
- 382 Yim, J., Trippe, B. L., Bortoli, V. D., Mathieu, E., Doucet,
383 A., Barzilay, R., and Jaakkola, T. SE(3) diffusion model
384 with application to protein backbone generation, 2023.
385 URL <https://arxiv.org/abs/2302.02277>.

385

Appendix

386

A. Glossary

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

Table 2. Glossary of notation and terms used in the methods section.

Symbol	Description
Molecular Representation and Coordinates	
\mathcal{G}	Macrocycle graph, where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
\mathcal{V}	Set of atom vertices in a macrocycle graph \mathcal{G} .
\mathcal{E}	Set of edges (bonds) between the atoms in a macrocycle graph \mathcal{G} .
\mathcal{C}	Set (ensemble) of conformers for a macrocycle, where $\mathcal{C} = \{c_1, c_2, \dots, c_K\}$.
\mathcal{D}	Bond distances in a conformer ensemble \mathcal{C} .
Θ	Bond angles in a conformer ensemble \mathcal{C} .
\mathcal{T}	Dihedral (torsional) angles in a conformer ensemble \mathcal{C} .
ξ	Vector of all Cartesian coordinates in a conformer.
\mathbf{d}	Vector of all bond distances in a conformer.
θ	Vector of all bond angles in a conformer.
τ	Vector of all dihedral (torsional) angles in a conformer.
$d_{i,j}$	Bond distance between atoms v_i and v_j .
$\theta_{i,j,k}$	Bond angle between atoms v_i , v_j , and v_k .
$\tau_{i,j,k,l}$	Dihedral (torsional) angle between atoms v_i , v_j , v_k , and v_l .
ϕ	Dihedral angle of bond between nitrogen and α -carbon.
ψ	Dihedral angle of bond between α -carbon and carbonyl-carbon.
ω	Dihedral angle of bond between carbonyl-carbon and nitrogen (peptide bond).
Encoder Model	
\mathbf{a}_i	Atom features for vertex v_i .
θ_i	Bond angle corresponding to vertex v_i .
τ_i	Dihedral (torsional) angle corresponding to vertex v_i .
\mathbf{x}_i	Internal coordinates corresponding to vertex v_i , where $\mathbf{x}_i = [\theta_i, \tau_i]$.
\mathbf{v}_i	Input/hidden representation for vertex v_i .
\mathbf{z}_i	Self-attention output for vertex v_i .
α_{ij}	Attention probability between vertices v_i and v_j .
e_{ij}	Unnormalized attention score between vertices v_i and v_j .
d_z	Attention head dimensionality.
\mathbf{P}_{ij}^K	Cyclic relative positional embedding between vertices v_i and v_j .
\mathbf{W}^K	Key matrix.
\mathbf{W}^Q	Query matrix.
\mathbf{W}^V	Value matrix.
\mathbf{W}^D	Graph-distance embedding matrix.
Diffusion Process	
\mathbf{x}_t	Noised internal coordinates (bond angle and torsion) at time step t .
$q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$	Forward transition probability.
$p_\Xi(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$	Diffusion model (reverse transition probability) parameterized by Ξ .
β_t	Variance from cosine variance schedule at time step t .
ϵ_t	Noise scale at time step t .
Miscellaneous	
$\hat{\cdot}$	Denotes predicted/generated quantity.
$w(\cdot)$	Function to wrap within $[-\pi, \pi]$ range, $w(\tau) := (\tau + \pi) \bmod (2\pi) - \pi$.
δ	Threshold for evaluating Coverage metric.

B. Dataset Description

Table 3. Dataset statistics for CREMP (Grambow et al., 2023).

Residues	Molecules	Conformers				
		Count	Mean	Median	Std. Dev.	Min.
4	17,842	12,205,128	684	508	677	1
5	13,644	14,134,609	1,036	825	824	6
6	4,712	4,921,068	1,044	879	764	28
Total	36,198	31,260,805	864	656	768	1
						12,268

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). 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).

C. Training Details

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} .

We use the discrete-time diffusion model from Wu et al. (2022) that formulates the forward transition probability using a wrapped normal distribution,

$$\begin{aligned} q(\mathbf{x}_t | \mathbf{x}_{t-1}) &= \mathcal{N}_{\text{wrapped}} \left(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I} \right) \\ &= \frac{1}{\beta_t \sqrt{2\pi}} \sum_{\mathbf{k} \in \mathbb{Z}^n} \exp \left(\frac{-\|\mathbf{x}_t - \sqrt{1 - \beta_t} \mathbf{x}_{t-1} + 2\pi\mathbf{k}\|^2}{2\beta_t^2} \right) \end{aligned} \quad (1)$$

instead of a standard normal distribution (Jing et al., 2022), where \mathbf{x}_t represents the noised internal coordinates (bond angle and torsion) at time step t . The diffusion model, $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t)$, parameterized by Ξ , reverses the process to denoise a wrapped normal distribution toward the data distribution. In the conditional setting, we further guide the diffusion process by learning $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathcal{G})$ in order to draw samples from the ensemble for a specific macrocycle, \mathcal{G} . We use the same cosine variance schedule as Wu et al. (2022) and Nichol & Dhariwal (2021) for $\beta_t \in (0, 1)_{t=1}^T$, but with significantly fewer time steps (typically, $T = 20$). $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t)$ and $p_{\Xi}(\mathbf{x}_{t-1} | \mathbf{x}_t, \mathcal{G})$ are trained using the simplified objective from Ho et al. (2020) to train a neural network, $\epsilon_{\Xi}(\mathbf{x}_t, t)$, to predict the noise present at a given time step by minimizing a smooth L1 loss (Girshick, 2015) wrapped by $w(\mathbf{x}) = (\mathbf{x} + \pi) \bmod (2\pi) - \pi$:

$$\begin{aligned} \mathbf{d}_w &= w \left(\epsilon - \epsilon_{\Xi} \left(w \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon \right), t \right) \right) \\ L_w &= \frac{1}{N} \sum_{i=1}^N \begin{cases} 0.5 \frac{d_{w,i}^2}{\beta_L} & \text{if } |d_{w,i}| < \beta_L \\ |d_{w,i}| - 0.5\beta_L & \text{otherwise} \end{cases} \end{aligned} \quad (2)$$

with $\beta_L = 0.1\pi$ as the transition point between L1 and L2 regimes (Wu et al., 2022), $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. We sample time steps uniformly from $t \sim U(0, T)$ during training and shift the bond angles and dihedrals using the element-wise means from the training data.

495 Cyclic peptides with only standard peptide bonds (i.e., homodetic peptides) do not have a natural starting residue and
 496 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
 497 is denoted by its one-letter code with “cyclo” indicating cyclization of the sequence. We therefore apply cyclic relative
 498 positional encodings:

$$500 \quad \mathbf{z}_i = \sum_{j=1}^n \alpha_{ij} (\mathbf{v}_j \mathbf{W}^V), \quad \text{where } \alpha_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^n \exp e_{ik}} \quad (3)$$

$$505 \quad e_{ij} = \frac{\mathbf{v}_i \mathbf{W}^Q (\mathbf{v}_j \mathbf{W}^K + \mathbf{p}_{ij}^K)^T}{\sqrt{d_z}} \quad \text{with } \mathbf{p}_{ij}^K = \underbrace{\mathbf{W}_{(i-j) \bmod n}^D}_{\text{forward}} + \underbrace{\mathbf{W}_{(i-j) \bmod (-n)}^D}_{\text{backward}} \quad (4)$$

509 These cyclic relative position representations encode bidirectional edge relationships between each atom by specifying
 510 forward and reverse distances in the macrocycle. The relative position of any neighboring atom is uniquely defined by its
 511 forward and reverse graph distances in the embedding lookup \mathbf{W}^D .

513 D. Sampling Details

515 We also use the sampling scheme from Wu et al. (2022). During inference, we first sample \mathbf{x}_T from a wrapped normal
 516 distribution and iteratively generate \mathbf{x}_0 from $t = T$ to $t = 1$ using

$$519 \quad \mathbf{x}_{t-1} = w \left(\frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_{\Xi}(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{x} \right) \quad (5)$$

522 where $\sigma_t = \sqrt{\beta_t(1 - \bar{\alpha}_{t-1})/(1 - \bar{\alpha}_t)}$ is the variance of the reverse process and $\mathbf{z} = \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$ and $\mathbf{z} = \mathbf{0}$ otherwise.

524 E. Model Details and Hyperparameters

526 Our model is a BERT transformer (Devlin et al., 2019) with cyclic relative positional encodings. The model input is a
 527 sequence of internal coordinates (and atom features for the conditional model). We linearly upscale the two-dimensional
 528 model input (bond angles and dihedrals) and separately upscale the atom features. Angles and atom features are then
 529 concatenated. The time step is embedded using random Fourier embeddings (Song et al., 2021) and added to the upscaled
 530 input. The combined embeddings are passed through the BERT transformer, the output of which is passed through a
 531 two-layer feed-forward network with GELU activation and layer normalization. Relevant hyperparameters are shown in
 532 Table 4.

535 Table 4. Hyperparameters.

536 Parameter	537 Value
538 Angle embedding size	256
539 Atom feature embedding size	128
540 Encoder layer dimensionality (hidden size)	384
541 Number of hidden layers	12
542 Number of attention heads	12
543 Feed-forward layer dimensionality (intermediate size)	512
544 Optimizer	AdamW
545 Learning rate	10^{-3}
546 Maximum number of epochs	1000
547 Warmup epochs	10
548 Batch size	8192

To condition on the atom sequence, we encode each atom using features of the atom itself and a Morgan fingerprint representation of the side chain attached to the atom (including the atom itself). The atom features include the atomic number, a chiral tag (L, D, or no chirality), aromaticity, hybridization, degree, valence, number of hydrogens, charge, sizes of rings that the atom is in, and the number of rings that the atom is in. The Morgan fingerprint is a count fingerprint with radius 3 and size 32.

F. Optimization for Back Conversion to Cartesian Ring Coordinates

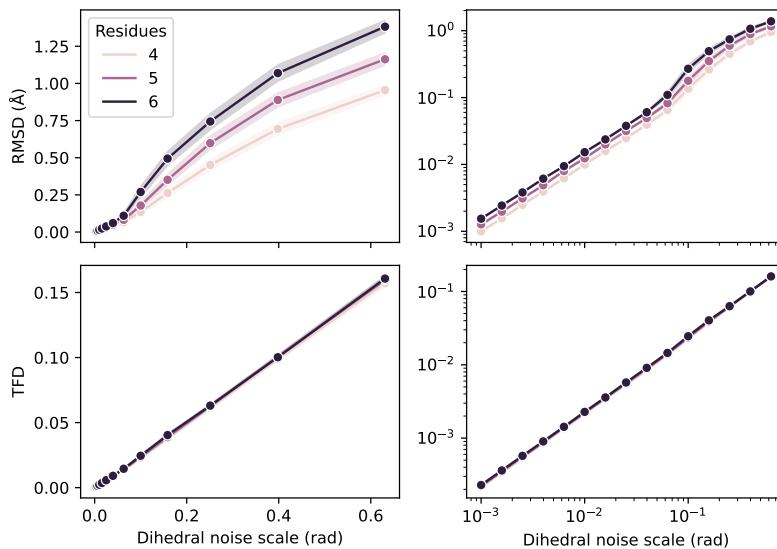


Figure 4. Our constrained optimization procedure is robust to noise as illustrated by a synthetic test of applying noise to the dihedral angles, recovering Cartesian coordinates using Equation (6), and comparing to the initial geometry in terms of RMSD and TFD.

To convert from the set of redundant internal coordinates predicted by the model back to Cartesian coordinates, we solve the following optimization to obtain a set of Cartesian coordinates that exactly satisfies the known bond distances in the ring.

$$\hat{\xi} = \arg \min_{\xi} \|\theta(\xi) - \hat{\theta}\|^2 + \|w(\tau(\xi) - \hat{\tau})\|^2 \quad \text{subject to: } d(\xi) = d_{\text{true}} \quad (6)$$

Here, we find the set of Cartesian coordinates, $\hat{\xi}$, that minimize the squared error against the internal coordinates $\hat{\theta}$ and $\hat{\tau}$ sampled by the diffusion process while satisfying bond distance equality constraints using known bond distances, d_{true} , from the training data.

To demonstrate that this procedure is robust to noise, we repeatedly embed 4-, 5-, and 6-mer backbones in 3D using RDKit distance geometry, extract their (redundant) internal coordinates, and add noise to the dihedral angles at different noise scales (standard deviation of a normal distribution) while ensuring that angles always remain in the $[-\pi, \pi]$ range. This creates a set of inconsistent, redundant dihedral angles, i.e., there exists no direct correspondence in Cartesian coordinates. We recover a possible Cartesian configuration using Equation (6) and compute RMSD and TFD for the ring atoms compared to the “true” internal coordinates from the RDKit geometry. Figure 4 shows that even moderate errors (~ 0.1 rad) result in very small errors in terms of both RMSD (~ 0.1 Å) and TFD (~ 0.02).

Notably, the optimization problem in Equation (6) is non-convex and requires a suitable initial guess to perform well. We assign this initial guess by obtaining a Cartesian geometry using the approach of sequentially setting atom positions according to the sequence of bond distances, angles, and torsions starting from one of the atoms in the ring. The starting atom is selected such that the redundant bond distance most closely matches the true bond distance (obtained from the training data). Improving the initial guess could be another direction for future research.

605 G. Software

606 All experiments were performed using Python and standard numerical libraries. For cheminformatics analysis, all molecules
 607 were processed using either OpenEye Applications and Toolkits (Software, 2022) or the open-source cheminformatics library
 608 RDKit (Landrum, 2006). We implemented all experiments in Python using PyTorch (Paszke et al., 2019) and PyTorch
 609 Lightning (Falcon & The PyTorch Lightning team, 2019). Transformers were implemented using BERT models within
 610 HuggingFace Transformers (Wolf et al., 2020).

612 H. Hardware

613 Each model was trained on a single NVIDIA A100 GPU with 80 GB VRAM using 12 CPUs for data loading and 96 GB of
 614 memory.

617 I. Evaluation

618 For unconditional generation, we use Kullback-Leibler divergence to measure the difference in sample quality. For
 619 unconditional generation, we evaluate the quality of our generated macrocycle backbones using root-mean-squared-deviation
 620 (RMSD) between backbone atom coordinates, similar to previous work on small-molecule conformer generation. We use
 621 several metrics including **Matching** and **Coverage** (Xu et al., 2021; Ganea et al., 2021; Jing et al., 2022), and for each we
 622 report recall and precision. We note that although RMSD is widely used to assess conformer quality, its utility for comparing
 623 backbones is more limited, as sampled backbones with highly unrealistic or energetically unfavorable torsions can exhibit
 624 low RMSD values. We additionally report the torsion fingerprint deviation (TFD) (Schulz-Gasch et al., 2012; Wang et al.,
 625 2020) to evaluate the quality of the torsional profiles. RMSD provides a measure of distance between two conformers based
 626 on a least-squares alignment of their respective atomic positions, while TFD gives a normalized measure of matched torsion
 627 angles between backbone geometries.

628 To measure both diversity and quality of the generated ensembles, we follow previous work and leverage four RMSD-based
 629 metrics (Xu et al., 2021; Ganea et al., 2021) with the difference that we only evaluate RMSD on macrocycle atoms. The
 630 *recall-based* **Coverage** metric measures the percentage of correctly generated conformers at a certain RMSD threshold,
 631 δ_{RMSD} . For a ground-truth ensemble \mathcal{C} and a generated ensemble $\hat{\mathcal{C}}$:

$$632 \text{RMSD-COV-R}(\hat{\mathcal{C}}, \mathcal{C}) = \frac{1}{|\mathcal{C}|} \left| \left\{ c \in \mathcal{C} : \exists \hat{c} \in \hat{\mathcal{C}}, \text{RMSD}(\hat{c}, c) \leq \delta_{\text{RMSD}} \right\} \right| \quad (7)$$

633 The *recall-based* **Matching** metric measures the average RMSD across the closest-matching (minimum-RMSD) generated
 634 conformer for each ground-truth conformer:

$$635 \text{RMSD-MAT-R}(\hat{\mathcal{C}}, \mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \min_{\hat{c} \in \hat{\mathcal{C}}} \text{RMSD}(\hat{c}, c) \quad (8)$$

636 The other two RMSD-based metrics are *precision* metrics that are defined identically, except that the ground-truth and
 637 generated ensembles are switched, and therefore constitute a measure of how many generated conformers are of high quality.

638 Analogous to the RMSD-based metrics, we define four metrics based on torsion fingerprint deviation (TFD) (Schulz-Gasch
 639 et al., 2012; Wang et al., 2020) to measure diversity and quality in terms of the torsional profiles of the generated rings:

$$640 \text{TFD-COV-R}(\hat{\mathcal{C}}, \mathcal{C}) = \frac{1}{|\mathcal{C}|} \left| \left\{ c \in \mathcal{C} : \exists \hat{c} \in \hat{\mathcal{C}}, \text{TFD}(\hat{c}, c) \leq \delta_{\text{TFD}} \right\} \right| \quad (9)$$

$$641 \text{TFD-MAT-R}(\hat{\mathcal{C}}, \mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \min_{\hat{c} \in \hat{\mathcal{C}}} \text{TFD}(\hat{c}, c) \quad (10)$$

642 TFD quantifies how well the macrocycle torsion angles match between two conformers and is given by (Wang et al., 2020):

643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659

$$660 \quad 661 \quad \text{TFD}(\hat{c}, c) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\pi} |w(\tau_i(\hat{c}) - \tau_i(c))| \quad 662 \quad 663 \quad 664 \quad 665 \quad 666 \quad 667 \quad 668 \quad 669 \quad 670 \quad 671 \quad 672 \quad 673 \quad 674 \quad 675 \quad 676 \quad 677 \quad 678 \quad 679 \quad 680 \quad 681 \quad 682 \quad 683 \quad 684 \quad 685 \quad 686 \quad 687 \quad 688 \quad 689 \quad 690 \quad 691 \quad 692 \quad 693 \quad 694 \quad 695 \quad 696 \quad 697 \quad 698 \quad 699 \quad 700 \quad 701 \quad 702 \quad 703 \quad 704 \quad 705 \quad 706 \quad 707 \quad 708 \quad 709 \quad 710 \quad 711 \quad 712 \quad 713 \quad 714$$

$$\text{TFD}(\hat{c}, c) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\pi} |w(\tau_i(\hat{c}) - \tau_i(c))| \quad (11)$$

where $\tau_i(c)$ extracts the i -th macrocycle torsion angle of conformer c and $w(\cdot)$ ensures that the deviation is wrapped correctly around the $[-\pi, \pi]$ boundary. Each torsion deviation is normalized by the maximum (absolute) deviation, π , so that TFD lies in $[0, 1]$.

J. Conformer Generation Baselines

RDKit ETKDGv3 RDKit baselines used ETKDGv3 (Riniker & Landrum, 2015; Wang et al., 2020) with macrocycle torsion preferences. We first embedded up to $2K$ conformers (where K is the number of true conformers) using `EmbedMultipleConfs` with random coordinate initialization (`useRandomCoords=True`), which has been shown to be beneficial for generating macrocycle geometries (Wang et al., 2020). Conformers were subsequently optimized using MMFF94 (Halgren, 1996) as implemented in RDKit and sorted by energy. Finally, the sorted conformers were filtered based on heavy-atom RMSD with a threshold of 0.1 \AA .

OpenEye OMEGA: Macrocyclic Mode OMEGA baselines were performed using OpenEye Applications (2022.1.1) with OMEGA (v. 4.2.0) (Hawkins et al., 2010; Hawkins & Nicholls, 2012) in macrocycle mode (Spellmeyer et al., 1997). Conformational ensembles were generated with the following macrocycle settings: `maxconfs=2K`, `ewindow=20`, `rms=0.1`, `dielectric_constant=5.0`, where K corresponds to the number of ground truth conformers from the original CREST ensemble in the CREMP dataset. The dielectric constant was set to 5.0 (chloroform) to most closely mimic the implicit chloroform solvation used in CREMP.

GeoDiff We used the original paper implementation of GeoDiff from Xu et al. (2022) available at <https://github.com/MinkaiXu/GeoDiff>, which we retrained to convergence on the CREMP dataset with the same data splits. As with all the other methods, we evaluated GeoDiff by sampling $2K$ conformers for each molecule. Inference for GeoDiff uses 5,000 time steps, which required more than 24 h for all test set molecules on 20 A100 GPUs.

K. Additional Unconditional Generation Results

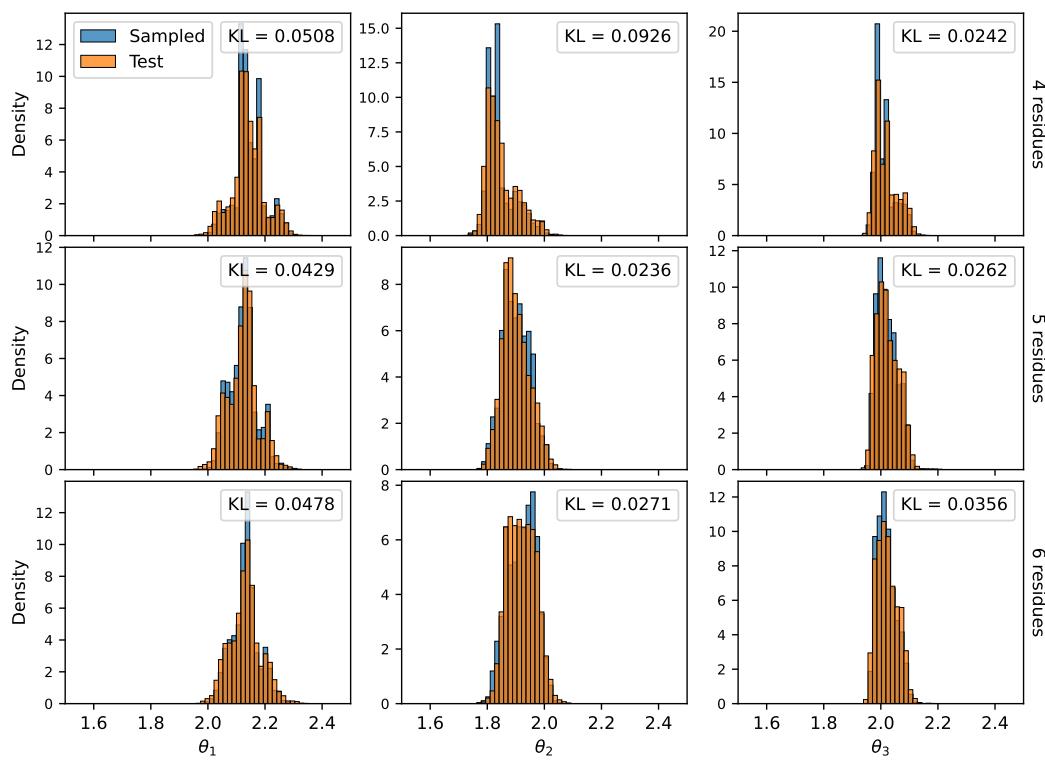


Figure 5. Bond angle distributions split by number of residues.

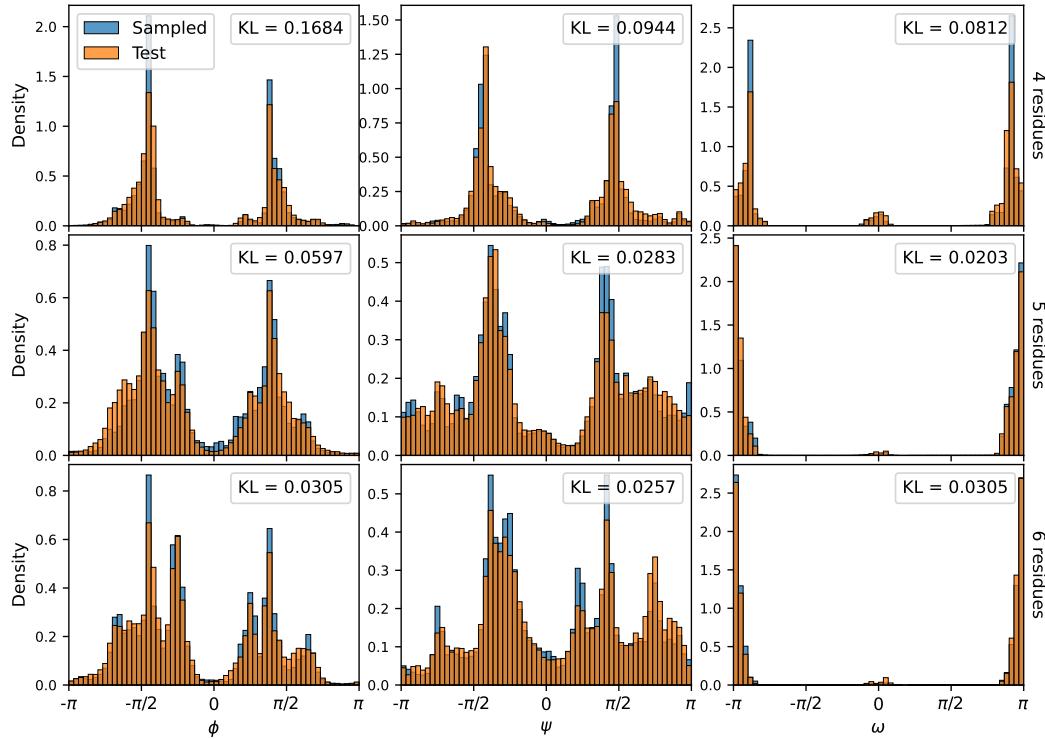


Figure 6. Dihedral angle distributions split by number of residues.

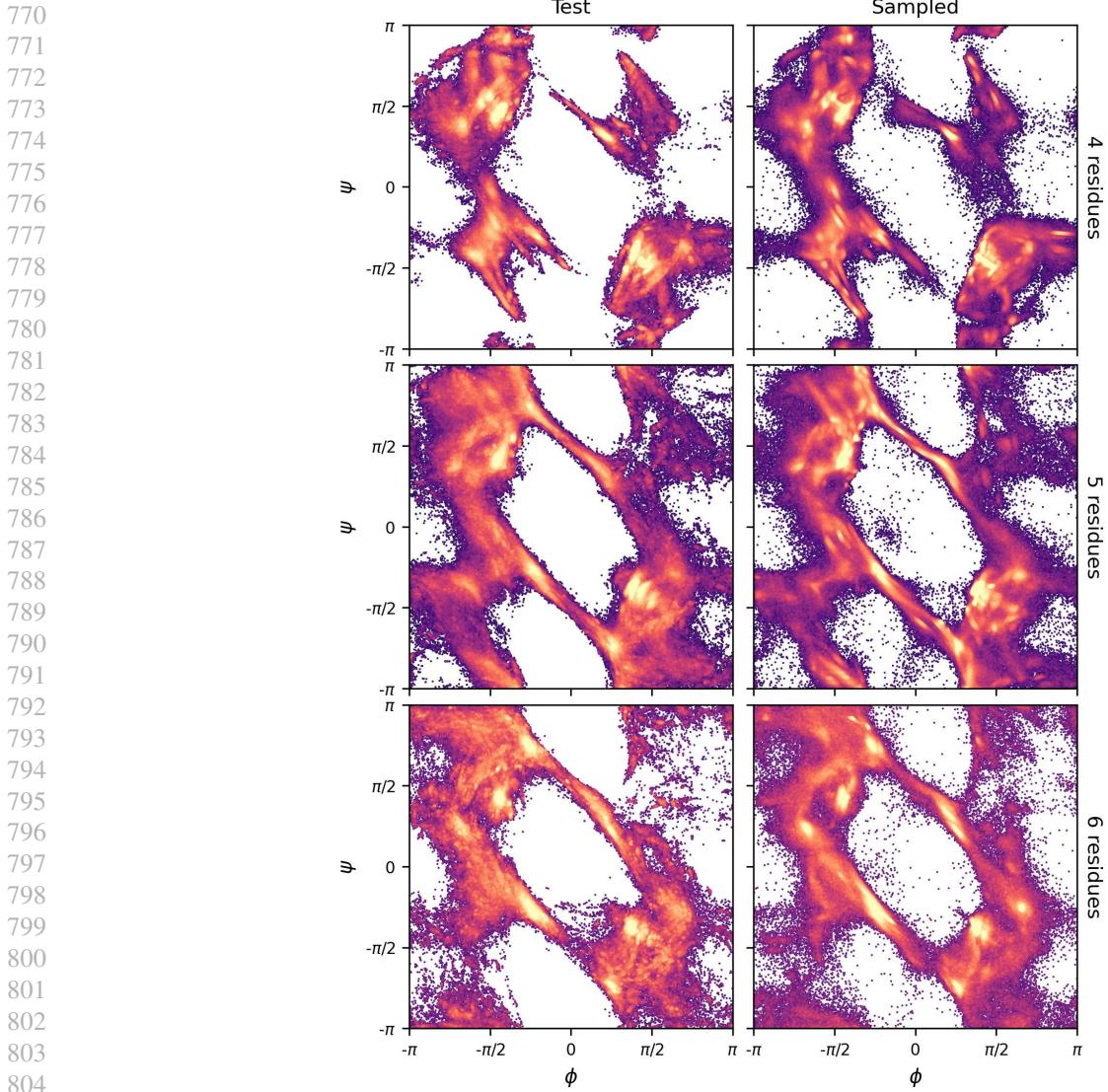


Figure 7. Ramachandran distributions split by number of residues.

L. Additional Sequence-Conditioned Generation Results

L.1. Distributions

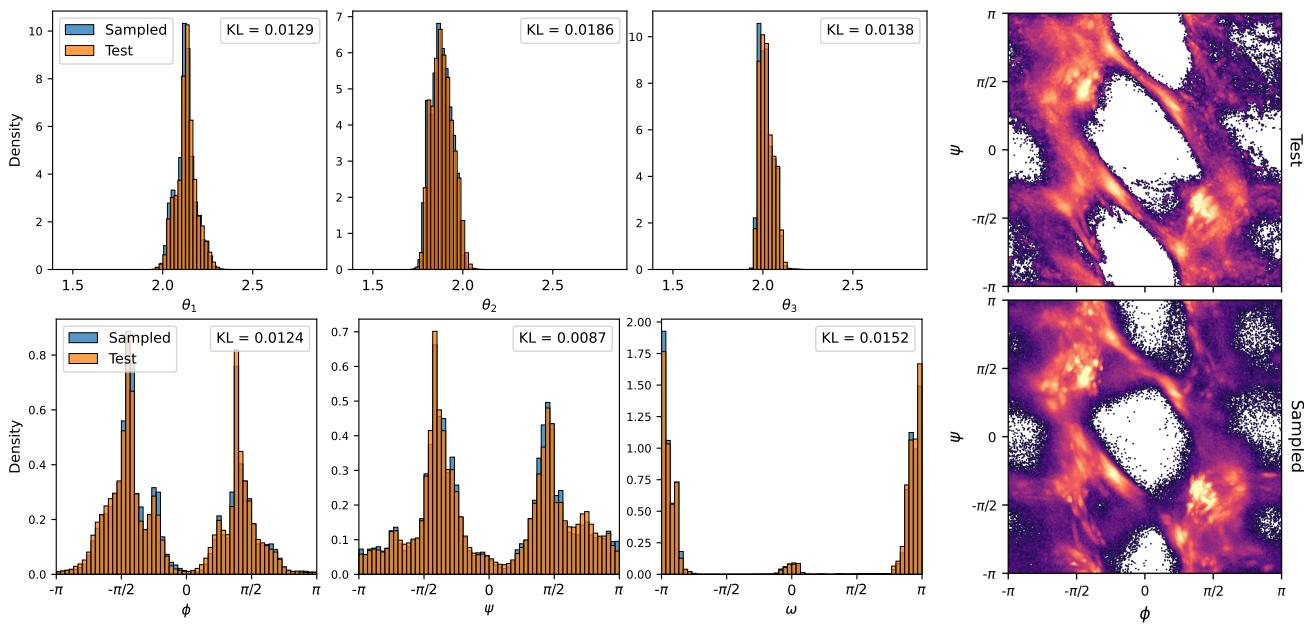


Figure 8. Comparison of the bond angle and dihedral distributions from the held-out test set and in the *conditionally* generated samples.

Figure 9 shows the Ramachandran plots split by number of residues for the conditional model and illustrates the effect of Equation (6) to reconstruct realizable Cartesian geometries from the set of redundant internal coordinates predicted by the model. Notably, while the reconstructed geometries still reproduce the joint distribution over dihedral angles well, several artifacts are introduced as a result of the optimization, which motivates the further development of generative methods that directly incorporate the cyclic constraints into the diffusion process itself.

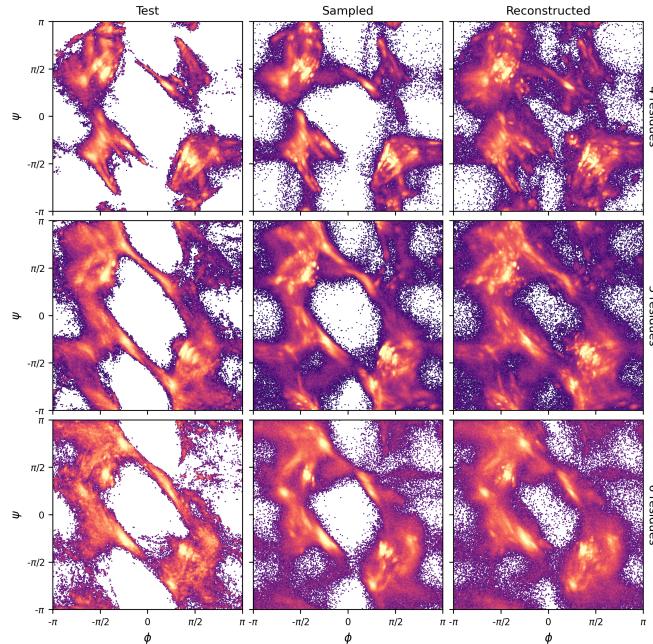


Figure 9. Ramachandran distributions for *conditionally* generated samples split by number of residues. The “Reconstructed” column shows the distributions after converting to Cartesian coordinates using the SLSQP optimization in Equation (6).

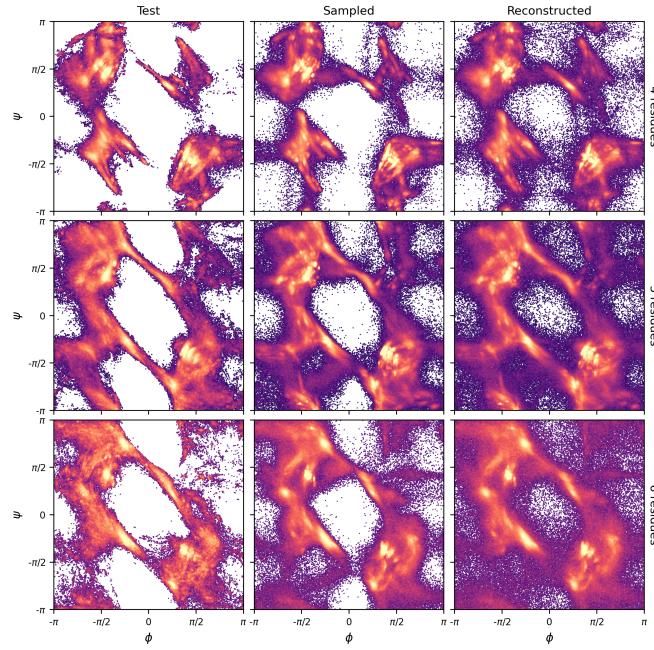


Figure 10. Ramachandran distributions for *conditionally* generated samples split by number of residues. The “Reconstructed” column shows the distributions after converting to Cartesian coordinates using the *naive* approach of setting internal coordinates sequentially from the starting atom that results in the structure with the smallest error in the last, redundant bond distance.

L.2. 3D Backbone Generation

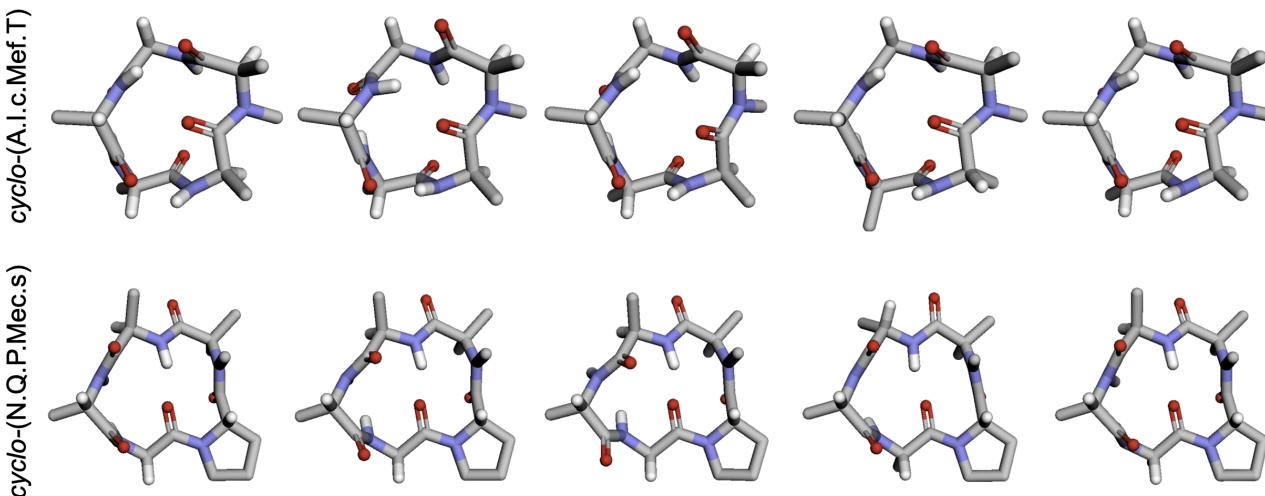


Figure 11. RINGER accurately generates the 3D backbones of the lowest-energy reference structure and the closest matching conformer (based on ring-atom RMSD) from each method. For RINGER: C from N-methyl, O from carbonyl, hydrogens, and proline carbons were inferred via MMFF94 optimization constraining the backbone geometry. Corresponding Ramachandran plots are shown in Figure 3.

L.3. Coverage

Figure 12 shows that RINGER outperforms all baselines over a wide range of thresholds used for evaluating Coverage. The plateau in RMSD precision (but not for TFD) of RINGER with post-processing is a result of the optimization converging to unrealistic geometries that nonetheless match the true torsions well. This motivates further development of methods to natively handle the cycle constraints as a future direction.

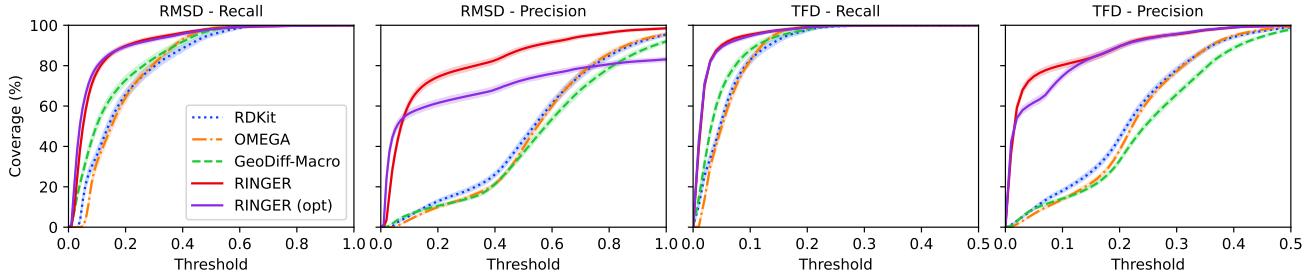


Figure 12. Comparison of mean coverage when varying the threshold across RMSD (left) and TFD (right). Translucent error bands correspond to 95% confidence intervals.

L.4. Confidence Intervals

Table 5. Performance metrics for sequence-conditioned generation of macrocycles using RMSD of macrocycle ring atoms. Coverage is evaluated at a threshold of 0.1 Å. Metrics shown as 95% bootstrap confidence intervals.

Method	RMSD – Recall		RMSD – Precision	
	Coverage (%) ↑	MAT (Å) ↓	Coverage (%) ↑	MAT (Å) ↓
RDKit (Wang et al., 2020)	(33.4, 38.4)	(5.9, 12.8)	(0.179, 0.195)	(0.148, 0.168)
OMEGA (Software, 2022)	(29.9, 34.6)	(4.2, 10.2)	(0.18, 0.193)	(0.148, 0.174)
GeoDiff-Macro (Xu et al., 2022)	(48.5, 53.2)	(47.1, 58.9)	(0.144, 0.159)	(0.111, 0.129)
RINGER	(75.5, 78.4)	(82.3, 86.2)	(0.087, 0.096)	(0.069, 0.075)
RINGER (opt)	(78.3, 80.9)	(84.3, 88.0)	(0.08, 0.09)	(0.062, 0.068)

Method	TFD – Recall		TFD – Precision	
	Coverage ↑	MAT ↓	Coverage ↑	MAT ↓
RDKit (Wang et al., 2020)	(5.1, 6.1)	(0.6, 1.3)	(0.529, 0.552)	(0.492, 0.515)
OMEGA (Software, 2022)	(3.4, 4.0)	(0.9, 1.7)	(0.548, 0.567)	(0.517, 0.535)
GeoDiff-Macro (Xu et al., 2022)	(5.9, 6.9)	(2.3, 3.6)	(0.581, 0.604)	(0.546, 0.571)
RINGER	(59.4, 63.1)	(66.0, 72.0)	(0.175, 0.196)	(0.114, 0.129)
RINGER (opt)	(54.6, 58.1)	(58.2, 65.5)	(0.421, 0.462)	(0.321, 0.384)

Table 6. Performance metrics for sequence-conditioned generation of macrocycles using torsional fingerprint deviation (TFD) of ring torsions. Coverage is evaluated at a threshold of 0.05. Metrics shown as 95% bootstrap confidence intervals.

Method	TFD – Recall		TFD – Precision	
	Coverage ↑	MAT ↓	Coverage ↑	MAT ↓
RDKit (Wang et al., 2020)	(50.5, 55.3)	(48.6, 63.2)	(0.057, 0.062)	(0.048, 0.054)
OMEGA (Software, 2022)	(47.4, 52.0)	(40.2, 56.5)	(0.059, 0.063)	(0.052, 0.059)
GeoDiff-Macro (Xu et al., 2022)	(66.0, 70.2)	(79.1, 86.1)	(0.046, 0.051)	(0.035, 0.039)
RINGER	(89.1, 91.0)	(94.3, 95.6)	(0.023, 0.026)	(0.018, 0.02)
RINGER (opt)	(88.2, 90.1)	(93.7, 95.0)	(0.023, 0.026)	(0.018, 0.02)

990 **L.5. Additional Evaluation**

991
992 *Table 7.* Evaluating RINGER (opt) trained and sampled with different numbers of timesteps.

993 994 995 996 997 998 999 1000	RMSD – Recall				RMSD – Precision			
	Timesteps	Coverage (%) ↑	MAT (Å) ↓	Coverage (%) ↑	MAT (Å) ↓	Mean	Med.	Mean
20	79.7	86.3	0.084	0.065	56.4	62.7	0.441	0.356
50	80.8	88.0	0.082	0.061	60.5	68.9	0.431	0.335
100	81.5	88.9	0.080	0.060	58.0	65.2	0.443	0.365

1001
1002 Table 8 shows that bond angles are required in addition to dihedral angles in order for the model to perform well. To
1003 reconstruct Cartesian geometries using the dihedral-only model, we modified Equation (6) to include inequality constraints
1004 for the bond angles where the upper and lower limit are determined by the standard deviations of bond angles from the
1005 training data.

1006
1007 *Table 8.* Evaluating RINGER (opt) trained only with dihedral angles.

1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044	RMSD – Recall				RMSD – Precision			
	Coverage (%) ↑		MAT (Å) ↓		Coverage (%) ↑		MAT (Å) ↓	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
$\mathbf{x}_i = [\theta_i, \tau_i]$	79.7	86.3	0.084	0.065	56.4	62.7	0.441	0.356
$\mathbf{x}_i = [\tau_i]$	66.8	73.5	0.130	0.101	41.1	37.6	0.469	0.417

1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044	TFD – Recall				TFD – Precision			
	Coverage (%) ↑		MAT ↓		Coverage (%) ↑		MAT ↓	
	Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
$\mathbf{x}_i = [\theta_i, \tau_i]$	89.2	94.3	0.024	0.019	61.8	68.9	0.068	0.044
$\mathbf{x}_i = [\tau_i]$	83.2	91.0	0.035	0.024	49.2	49.3	0.144	0.114