Do Deep Learning Methods Really Perform Better in Molecular Conformation Generation?

Gengmo Zhou[1,2]***Zhifeng Gao**[1]***Zhewei Wei**[2]**, Hang Zheng**[1]**, Guolin Ke**[1]†
[1]DP Technology, China
[2]Renmin University of China, China
`{zhougm, gaozf}@dp.tech`
`zhewei@ruc.edu.cn`
`{zhengh, kegl}@dp.tech`

Abstract

Molecular conformation generation (MCG) is a significant and vital problem in drug development. While traditional methods, like systematic searching, model-building, random searching, distance geometry, molecular dynamics, and Monte Carlo methods, have been developed to solve the MCG problem, they have limitations that are dependent on the molecular structures. Recently, deep learning-based MCG methods have been introduced, claiming to outperform the traditional methods. Surprisingly, we have developed a simple and cheap algorithm (parameter-free) based on traditional methods and found that it is comparable to, or even outperforms, deep learning-based MCG methods on widely used GEOM-QM9 and GEOM-Drugs benchmarks. Our algorithm design involves clustering of the RDKit-generated conformations. We believe that our findings can help the scientific community to revise deep learning methods and benchmarks for MCG. The code for the proposed algorithm is available at `https://gist.github.com/ZhouGengmo/5b565f51adafcd911c0bc115b2ef027c`.

1 Introduction

Molecular conformation generation (MCG) plays a crucial role in drug discovery [1; 2], as it aims to generate low-energy spatial arrangements of atoms in a molecule. MCG is closely related to various drug design tasks, including molecular property prediction [3; 4], protein-ligand complex [5; 6; 7], structure similarity [8; 9], and pharmacophore searching [10].

The traditional approach for MCG [11] involves using conformational search methods to generate a set of molecular conformations, followed by applying energy minimization methods to optimize them further using energy estimators, such as molecular mechanics (MM) or density functional theory (DFT) [12; 13]. Finally, a batch of optimized conformations with low energies are screened out based on their potential energy. Various traditional approaches have been proposed for different scenarios.

For instance, RDKit, a popular cheminformatics software in the chemical community, is widely used in MCG. RDKit uses ETKDG [14], a template-based method, in its default setting to generate conformations efficiently. Then, MM methods such as MMFF [15] or UFF [16] are applied to perform structure optimization and obtain better conformations. Although a typical and efficient MCG method, it suffers from low performance in sampling coverage due to the limitations of template-based sampling and in energy evaluation due to the low accuracy of MM force fields.

Recently, several works applied deep learning to MCG, aiming to generate widely distributed and low energy conformations efficiently. These works can be classified into two categories. One category is methods based on distance geometry [17; 18; 19; 20; 21], which first generates atoms' pair-distance

---

*Equal contribution.
†Corresponding authors.

matrix and then recoveries atoms' coordinates based on the matrix. Another category is methods that generate atoms' coordinates directly [22; 23; 24; 25]. With the advent of diffusion models, there are also diffusion based methods [25] proposed for MCG tasks.

In these deep learning models, GEOM[3] dataset is widely used for training and evaluation. It uses extensive sampling methods, based on the semi-empirical DFT, to generate 3D conformations. In practice[20], 40k molecules (200k conformations) are sampled from GEOM for training, and 200 molecules (about 20k conformations) are sampled for testing. As there are multiple conformations for one molecule, Coverage (COV) and Matching (MAT) are proposed to evaluate the performance of these models. In particular, COV is the coverage of the generated conformation to the reference one under a certain RMSD threshold; the larger, the better diversity. MAT measures the difference between the generated conformation and the reference one; the smaller, the better accuracy.

The deep learning models demonstrated superior performance compared to traditional methods in the aforementioned benchmark. However, despite the models being proposed several years ago, we could not identify any real-world applications utilizing these deep models for conformation generation. Consequently, we question whether these models truly surpass traditional approaches.

Remarkably, we develop a straightforward algorithm based on traditional techniques that outperforms most deep learning models. Specifically, we employed the EDKTG algorithm integrated in RDKit [26] and utilized three samplers: uniform, geometric, and energy samplers to generate diverse and low-energy conformations. We then applied an unsupervised clustering algorithm to obtain representative conformations.

Our unexpected discovery challenges the superiority of existing deep learning models. As a result, we reassess these methods and offer the following suggestions for future research.

- *The current benchmark may be wrong.* Our simple algorithm outperformed all previous deep models with ease, indicating that the current benchmark may be vulnerable to manipulation. We suspect that a more diverse range of conformations could achieve even better results, and have designed an additional experiment to explore this possibility explicitly.

- *The benchmark should focus on the end applications, not the standalone conformations.* Molecular conformations are highly flexible, varying significantly depending on the environment, and can be difficult to obtain through wet experiments. Additionally, different applications have different requirements for conformations. For example, in molecular docking [5; 6; 7], the goal is to find conformations that are close to the binding conformation of the target, while for quantum property prediction, the goal is to identify conformations with the lowest energy. Therefore, benchmarking the effect of conformation generation should be based on the specific requirements of downstream applications, or on the final performance of these applications.

2 RELATED WORK

For MCG, there have been many classical methods in computational chemistry. Recently, with the development of deep learning, researchers have also proposed some data-driven solutions. In the following, we provide a brief overview of the related work.

2.1 CLASSICAL METHODS

The traditional paradigm of MCG [11] is to first use the conformational search methods to generate a bunch of molecular conformations, and after the energy minimization (a conformation optimization method based on energy) of each conformation, a batch of molecular conformations with low energies are screened out based on the potential energy evaluated by density functional theory (DFT) [12; 27], semi-empirical DFT [28], or molecular mechanism (MM) [15; 16].

The core problem of MCG is the conformational search problem. Since there are many rotatable dihedral angles in each molecule, a small change of each rotatable dihedral angle will generate a new molecular conformation. This is a combinatorial explosion problem. It is an urgent problem to generate conformations with lower energy under limited resources. Several popular conformational search methods have been proposed. The system search method [29; 30] will uniformly sample the conformation of each variable dihedral angle at a certain step size, but it is not suitable for molecules

with more dihedral angles and molecules containing large rings. The random search algorithm [31] randomly perturbs each dihedral angle, and then calculates the acceptance probability based on the Monte Carlo method to decide whether to keep the generated conformation. It has a long convergence time and has repeated visits. The model-building method [32] decomposes large molecules into molecular fragments, extracts the conformation of the molecular fragments from the structural database, and then splices the fragments to form the original molecular conformation. ETKDG[14] and OMEGA[33] both use this technique. It performs poorly on complex systems and molecules with strong intramolecular interactions. The distance geometry method [34] regards the molecular conformation as a series of points defined by distance parameters, and uses a geometric search algorithm to search for the best conformation, which is more useful when the distance information is known. Molecular dynamics simulation [35] allows molecules to generate conformations based on Newtonian mechanics under the constraints of molecular force fields, which can generate a large number of molecular conformations, but consumes more computing resources than the above methods.

After the conformational search problem is stated, the generated conformations still need to be evaluated and screened out. The commonly used methods are electronic structure method and force field methods. In terms of energy estimation, force field methods, such as MMFF[15] and UFF[16], are much less accurate than electronic structure methods. However, quantum mechanical or density functional theory (DFT) methods are time-consuming and unacceptable in the face of large numbers of molecules.

2.2 DEEP LEARNING METHODS

Recently, with the development of deep learning, researchers expect to deal with MCG with data-driven solutions, and on the GEOM benchmark, deep learning methods outperformed traditional methods such as EDKTG. In earlier work, CVGAE [22] used VAE to generate atomic coordinates directly. However, this model could not maintain translation and rotation equivariance of the generated conformations and performed poorly on the benchmark. To ensure such geometrical properties of the conformation, some later works use intermediate structures, such as interatomic distances or torsion angles, to generate conformations. GraphDG[17] and CGCF[18] use VAE and flow to generate distance matrices, respectively, and then use classical distance geometry techniques to generate conformations iteratively. ConfVAE[19] design an end-to-end framework to enhance model performance using bilevel optimization. However, minor errors in the predicted distance matrix can cause the generated 3D structures to be unreasonable, such as local structural conflicts that do not fit the triangular inequality. ConfGF[20] proposes to learn the gradient of coordinates to alleviate this problem. DGSM[21] takes the non-bonded local and long-range interactions into account when modeling. But they still rely on intermediate structural elements. The method of generating all the atom coordinates in the molecule directly, which avoids the intermediate geometric elements mentioned above, has yet to lose its attention. Similar to the rigid rotor approximation, GeoMol[23] does not consider unnecessary degrees of freedom and only predicts local 3D structures and torsion angles. DMCG[24] constantly refines the coordinates along model blocks and proposes a loss function that is invariant to the roto-translation of the atom coordinates and the permutation of symmetric atoms. With the advent of diffusion models, methods, such as GeoDiff [25] and Torsional Diffusion [36], apply this latest technique to the conformation generation task.

3 METHOD

In this paper, we present a straightforward approach utilizing RDKit[26] and clustering post-processing. Our method involves three samplers to produce a range of low-energy conformations that are diverse in structure, as depicted in Fig. 1. The first sampler employs a uniform distribution of dihedral angles in molecules, without taking energy estimation into account. The second sampler, RDKit's default conformation generation setting, uses the EDKTG geometric sampler to assemble conformations with specific molecular fragments based on knowledge and structural databases. Finally, the energy sampler generates more accurate energy estimations for better conformational optimization by utilizing the RDKit and MMFF force field optimization methods.
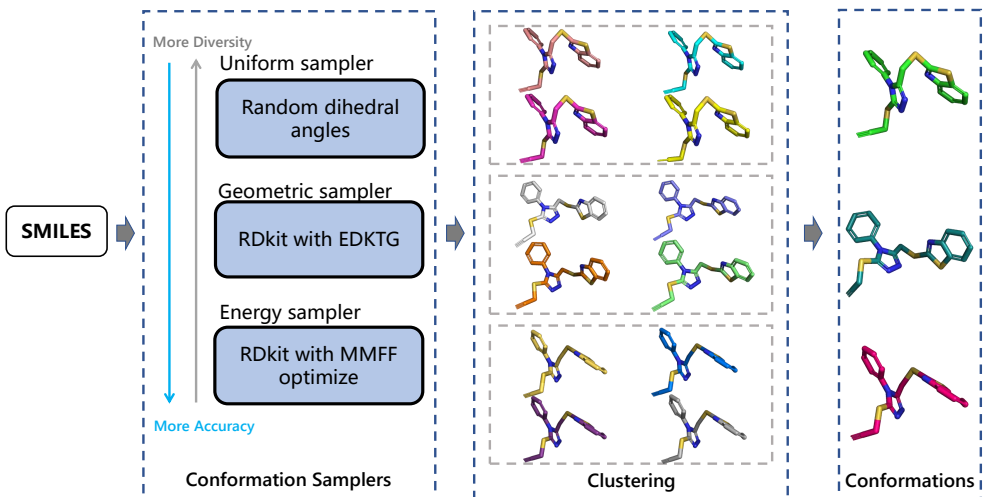
Figure 1: Overview of the proposed RDKit + Clustering algorithm.

By integrating those samplers, a simple unsupervised cluster algorithm is applied to select conformations with consideration of diversity and energy. It is noticed that the majority of clusters share conformations with lower energy, and different clusters share conformations with more diversity.

Specifically, our sampling process involves the uniform, geometric, and energy samplers, which are used in a ratio of 1:1:4, respectively. The number of energy samples, denoted by N_e, is determined by the formula $N_e = min(20N_{ref}, 2000)$, where N_{ref} refers to the number of reference conformations of the current molecule in the test set. The uniform and geometric samplers generate a quarter of the number of samples produced by the energy sampler, with N_u and N_g representing their respective sample sizes. Once a sufficient number of samples have been generated, we employ the K-means algorithm with $2N_{ref}$ clusters to cluster them, and select the center of each cluster as an output. Prior to clustering, we align and centralize the 3D coordinates of the samples using the kabsch algorithm [37], followed by flattening the vector of aligned coordinates for K-means. This method provides twice the number of conformations as N_{ref} for metric calculations, consistent with previous research [18; 20].

4 EXPERIMENT

Datasets and setup Following the settings in previous works [18; 20], we use GEOM-QM9 and GEOM-Drugs [3] dataset as the benchmark. The training set is constructed by randomly selecting 40,000 molecules from these two datasets, respectively, and each molecule is taken with 5 possible low-energy conformations sorted by energy. From the remaining data, 200 molecules are randomly selected, and the number of conformations per molecule is limited to between 50 and 500. 22,408 and 14,324 conformations are available in the test set for GEOM-QM9 and GEOM-Drugs, respectively. Since our method (RDKit+Clustering) does not require training, for the inference in the test set, we generate the same number of conformations (twice the number of labeled conformations) as previous works do. And we also use the same metrics, Coverage (COV) and Matching (MAT) as in previous works. Higher COV means better diversity, while lower MAT means higher accuracy. Despite the proposal of several new metrics and benchmarks for evaluation in previous works, the absence of a standardized metric, other than COV and MAT, limits our ability to make systematic comparisons on them. More details about metrics are shown in Appendix A.

Baselines We compare our method with 10 competitive baselines. RDKit [14], i.e., original EDKTG with MMFF, is a traditional conformation generation method based on distance geometry. The introduction of other baselines can be found in Sec. 2.2. A recent model known as torsional

Table 1: Performance on GEOM-QM9 and GEOM-Drugs. The best results are marked **bold**.

Dataset	QM9				Drugs			
	COV($\uparrow$, %)		MAT($\downarrow$, Å)		COV($\uparrow$, %)		MAT($\downarrow$, Å)	
Methods	Mean	Median	Mean	Median	Mean	Median	Mean	Median
RDKit	83.26	90.78	0.3447	0.2935	60.91	65.70	1.2026	1.1252
CVGAE	0.09	0.00	1.6713	1.6088	0.00	0.00	3.0702	2.9937
GraphDG	73.33	84.21	0.4245	0.3973	8.27	0.00	1.9722	1.9845
CGCF	78.05	82.48	0.4219	0.3900	53.96	57.06	1.2487	1.2247
ConfVAE	80.42	85.31	0.4066	0.3891	53.14	53.98	1.2392	1.2447
ConfGF	88.49	94.13	0.2673	0.2685	62.15	70.93	1.1629	1.1596
GeoMol	71.26	72.00	0.3731	0.3731	67.16	71.71	1.0875	1.0586
DGSM	91.49	95.92	0.2139	0.2137	78.73	94.39	1.0154	0.9980
GeoDiff	92.65	95.75	0.2016	0.2006	88.45	97.09	0.8651	0.8598
DMCG	94.98	98.47	0.2365	0.2312	**91.27**	**100.00**	0.8287	0.7908
RDKit + Clustering	**97.65**	**100.00**	**0.1902**	**0.1818**	87.93	**100.00**	**0.8086**	**0.7838**

diffusion has been introduced [36], but we have not compared it in our paper due to the following reasons. Firstly, the method employs different splits for training and test data, making an apple-to-apple comparison difficult. Moreover, torsional diffusion is a combination of conventional methods with deep learning, rather than a purely deep learning approach. It relies on RDKit-generated conformations as input, and modifies their torsional angles.

Results The results are shown in Table 1. We report the mean and median of COV and MAT on GEOM-QM9 and GEOM-Drugs datasets. ConfVAE [19]and DGSM [21] results are from their papers, respectively. GeoMol[23] results are from DMCG [24] paper, which DMCG reproduced after aligning the data split. We found the test sets used in GeoDiff [25] and DMCG [24] are slightly different from baselines (different filtering conditions), so we use the released model parameters of GeoDiff to reproduce the results on the same test set. For DMCG, as model parameters are not released, we use its open-source codes to reproduce from scratch. Other baseline results are from ConfGF's paper. As shown in Table 1, our method exceeds most existing baselines in COV and MAT metrics and achieves comparable performance to SOTA.

Ablation study As mentioned in Sec. 1, we hypothesize that more diverse conformations could easily achieve better results in GEOM. To demonstrate this explicitly, we design ablation studies for the sample size and sampler types. For the sample size, we set the energy sampler's sample size to twice, five times, and ten times the number of reference conformations N_{ref}. The sampling ratio of 3 samplers remains the same as before. For the sampler type, we remove one of the 3 samplers separately in the experiment. The sampling ratio also remains the same.

The results are shown in Table 2. From the results, the smaller the number of samples, the worse the performance, although there is some marginal utility. In terms of samplers, the energy sampler mainly improves the accuracy of the generated conformations while decreasing the diversity; the uniform sampler mainly improves the diversity while making the accuracy decrease; geometric samplers have a positive impact on both accuracy and diversity.

Discussion While our approach surpasses most of the baselines, we have doubts that the benchmark utilized can fulfill the needs of real-world conformation generation tasks. The ablation studies indicate that this benchmark can be improved by generating more diverse conformations. Furthermore, the requirements for conformations differ depending on the applications. For instance, in molecular docking, we require conformations that are similar to the binding conformation. The ensemble of bioactive conformations must be taken into account when using the conformation generation model in drug design, which is overlooked by GEOM. In contrast, for quantum property prediction, we require conformations that are close to the DFT-optimized conformations. The accuracy of GEOM's semi-empirical DFT is insufficient to meet these requirements. Therefore, we need to benchmark the impact of conformation generation according to the downstream applications' needs or their final performance.

Table 2: Ablation studies for the number of samples and sampler type on GEOM-QM9. The best results are marked **bold**.

Dataset	QM9			
	COV($\uparrow$, %)		MAT($\downarrow$, Å)	
Methods	Mean	Median	Mean	Median
$2N_{ref}$	91.23	96.87	0.2223	0.2171
$4N_{ref}$	95.19	99.58	0.1988	0.1901
$10N_{ref}$	96.49	100.00	0.1941	0.1857
w/o Uniform sampler	93.56	99.55	**0.1867**	**0.1643**
w/o Geometric sampler	97.42	100.00	0.1984	0.1930
w/o Energy sampler	**98.01**	100.00	0.2511	0.2434
RDKit + Clustering	97.65	**100.00**	0.1902	0.1818

5 CONCLUSION

Our study presents a simple algorithm that utilizes traditional approaches and has demonstrated superior performance compared to most deep learning models. As a result, we conducted an investigation into the current settings utilized in deep learning methods MCG, and found that the current benchmark for MCG may contain inaccuracies. We recommend that future research should focus on the end applications in various MCG-related downstream applications, given that benchmarking the performance of MCG alone is inherently challenging. We are confident that with a reasonable benchmark, deep learning-based methods can significantly contribute towards the development of effective MCG models for specific tasks such as bioactive conformation generation, general conformation optimization, and conformation scoring.

REFERENCES

[1] A Crum-Brown and TR Fraser. The connection of chemical constitution and physiological action. *Trans R Soc Edinb*, 25(1968-1969):257, 1865.

[2] Corwin Hansch and Toshio Fujita. p-σ-π analysis. a method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, 86(8):1616–1626, 1964.

[3] Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):1–14, 2022.

[4] Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, pages 1–8, 2022.

[5] Kunal Roy, Supratik Kar, and Rudra Narayan Das. *Understanding the basics of QSAR for applications in pharmaceutical sciences and risk assessment*. Academic press, 2015.

[6] Garrett M Morris, Ruth Huey, William Lindstrom, Michel F Sanner, Richard K Belew, David S Goodsell, and Arthur J Olson. Autodock4 and autodocktools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16):2785–2791, 2009.

[7] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

[8] Paul CD Hawkins, A Geoffrey Skillman, and Anthony Nicholls. Comparison of shape-matching and docking as virtual screening tools. *Journal of medicinal chemistry*, 50(1):74–82, 2007.

[9] Ambrish Roy and Jeffrey Skolnick. Ligsift: an open-source tool for ligand structural alignment and virtual screening. *Bioinformatics*, 31(4):539–544, 2015.

[10] Christof H Schwab. Conformations and 3d pharmacophore searching. *Drug Discovery Today: Technologies*, 7(4):e245–e253, 2010.

[11] Paul CD Hawkins. Conformation generation: the state of the art. *Journal of Chemical Information and Modeling*, 57(8):1747–1756, 2017.

[12] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.

[13] Kyle A Baseden and Jesse W Tye. Introduction to density functional theory: Calculations by hand on the helium atom. *Journal of chemical education*, 91(12):2116–2123, 2014.

[14] Sereina Riniker and Gregory A Landrum. Better informed distance geometry: using what we know to improve conformation generation. *Journal of chemical information and modeling*, 55(12):2562–2574, 2015.

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

[16] Anthony K Rappé, Carla J Casewit, KS Colwell, William A Goddard III, and W Mason Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society*, 114(25):10024–10035, 1992.

[17] Gregor Simm and Jose Miguel Hernandez-Lobato. A generative model for molecular distance geometry. In *International Conference on Machine Learning*, pages 8949–8958. PMLR, 2020.

[18] 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*, 2020.

[19] Minkai Xu, Wujie Wang, Shitong Luo, Chence Shi, Yoshua Bengio, Rafael Gomez-Bombarelli, and Jian Tang. An end-to-end framework for molecular conformation generation via bilevel programming. In *International Conference on Machine Learning*, pages 11537–11547. PMLR, 2021.

[20] Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *International Conference on Machine Learning*, pages 9558–9568. PMLR, 2021.

[21] Shitong Luo, Chence Shi, Minkai Xu, and Jian Tang. Predicting molecular conformation via dynamic graph score matching. *Advances in Neural Information Processing Systems*, 34, 2021.

[22] Elman Mansimov, Omar Mahmood, Seokho Kang, and Kyunghyun Cho. Molecular geometry prediction using a deep generative graph neural network. *Scientific reports*, 9(1):1–13, 2019.

[23] Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34, 2021.

[24] Jinhua Zhu, Yingce Xia, Chang Liu, Lijun Wu, Shufang Xie, Tong Wang, Yusong Wang, Wengang Zhou, Tao Qin, Houqiang Li, et al. Direct molecular conformation generation. *arXiv preprint arXiv:2202.01356*, 2022.

[25] 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.

[26] Greg Landrum et al. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling, 2013.

[27] Mariana Rossi, Sucismita Chutia, Matthias Scheffler, and Volker Blum. Validation challenge of density-functional theory for peptides example of ac-phe-ala5-lysh+. *The Journal of Physical Chemistry A*, 118(35):7349–7359, 2014.

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

[29] Denise D Beusen, EF Berkley Shands, SF Karasek, Garland R Marshall, and Richard A Dammkoehler. Systematic search in conformational analysis. *Journal of Molecular Structure: THEOCHEM*, 370(2-3):157–171, 1996.

[30] Nicolas Sauton, David Lagorce, Bruno O Villoutreix, and Maria A Miteva. Ms-dock: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC bioinformatics*, 9(1):1–12, 2008.

[31] Stephen R Wilson, Weili Cui, Jules W Moskowitz, and Kevin E Schmidt. Applications of simulated annealing to the conformational analysis of flexible molecules. *Journal of computational chemistry*, 12(3):342–349, 1991.

[32] Andrew Smellie, Robert Stanton, Randy Henne, and Steve Teig. Conformational analysis by intersection: Conan. *Journal of computational chemistry*, 24(1):10–20, 2003.

[33] Paul CD Hawkins, A Geoffrey Skillman, Gregory L Warren, Benjamin A Ellingson, and Matthew T Stahl. Conformer generation with omega: algorithm and validation using high quality structures from the protein databank and cambridge structural database. *Journal of chemical information and modeling*, 50(4):572–584, 2010.

[34] David C Spellmeyer, Alex K Wong, Michael J Bower, and Jeffrey M Blaney. Conformational analysis using distance geometry methods. *Journal of Molecular Graphics and Modelling*, 15(1):18–36, 1997.

[35] Hideki Tsujishita and Shuichi Hirono. Camdas: An automated conformational analysis system using molecular dynamics. *Journal of computer-aided molecular design*, 11:305–315, 1997.

[36] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi S. Jaakkola. Torsional diffusion for molecular conformer generation. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.

[37] Wolfgang Kabsch. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of applied crystallography*, 26(6):795–800, 1993.

A Metrics in GEOM Dataset

A.1 RMSD

Root of Mean Squared Deviations (RMSD) is widely used to evaluate the difference between conformations in molecules. Before computing RMSD, the generated conformation is first aligned with the reference one, and the function Φ aligns conformations by applying rotations and translations to them:

$$\text{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \min_{\Phi}(\frac{1}{n}\sum_{i=1}^{n}||\Phi(\boldsymbol{R}_i) - \hat{\boldsymbol{R}}_i||^2)^{\frac{1}{2}} \tag{1}$$

where $\boldsymbol{R}$ and $\hat{\boldsymbol{R}}$ are the generated and reference conformation, i is the i-th heavy atom, and n is the number of heavy atoms.

A.2 COV & MAT

Coverage (COV) and Matching (MAT) is used to evaluate the performance of the conformation generation model in GEOM. Higher COV means better diversity, while lower MAT means higher accuracy. Formally, COV and MAT are denoted as:

$$\text{COV}(S_g, S_r) = \frac{\left|\left\{\boldsymbol{R} \in S_r | \text{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) < \delta, \hat{\boldsymbol{R}} \in S_g\right\}\right|}{|S_r|} \tag{2}$$

$$\text{MAT}(S_g, S_r) = \frac{1}{|S_r|}\sum_{\boldsymbol{R} \in S_r}\min_{\hat{\boldsymbol{R}} \in S_g}\text{RMSD}(\boldsymbol{R}, \hat{\boldsymbol{R}}) \tag{3}$$

where S_g and S_r are the set of generated and reference conformations, respectively, and δ is a given RMSD threshold. Following previous work [18; 20], for GEOM-QM9, the threshold is 0.5Å, and for GEOM-Drugs, the threshold value is 1.25Å.