Von Mises Mixture Distributions for Molecular Conformation Generation

Kirk Swanson [1] **Jake Williams** [1] **Eric Jonas** [1]

Abstract

Molecules are frequently represented as graphs, but the underlying 3D molecular geometry (the locations of the atoms) ultimately determines most molecular properties. However, most molecules are not static and at room temperature adopt a wide variety of geometries or *conformations*. The resulting distribution on geometries $p(x)$ is known as the Boltzmann distribution, and many molecular properties are expectations computed under this distribution. Generating accurate samples from the Boltzmann distribution is therefore essential for computing these expectations accurately. Traditional sampling-based methods are computationally expensive, and most recent machine learning-based methods have focused on identifying *modes* in this distribution rather than generating true *samples*. Generating such samples requires capturing conformational variability, and it has been widely recognized that the majority of conformational variability in molecules arises from rotatable bonds. In this work, we present VonMisesNet, a new graph neural network that captures conformational variability via a variational approximation of rotatable bond torsion angles as a mixture of von Mises distributions. We demonstrate that VonMisesNet can generate conformations for arbitrary molecules in a way that is both physically accurate with respect to the Boltzmann distribution and orders of magnitude faster than existing sampling methods.

1. Introduction

Accurate prediction of molecular properties is an important task in computational chemistry. Many physical and chemical properties are dependent on a molecule's 3D conformations, and some properties, such as spectroscopic properties like nuclear magnetic resonance (NMR) chemical shifts, are measured on timescales considerably longer than the timescale of molecular motion. At room temperature, most molecules are not static and adopt a wide variety of conformations. These properties are therefore computed as expectations with respect to the probability density that governs molecular conformations, known as the Boltzmann distribution. This density is given by

$$p(x) = \frac{e^{-E(x)/k_B T}}{Z}, \tag{1}$$

where $x \in \mathbb{R}^{n\times 3}$ is the set of atomic coordinates that define a conformation over the n atoms, $E(x)$ is the potential energy, T is the temperature, Z is a normalizing constant, and k_B is the Boltzmann constant (Tuckerman, 2010). $E(x)$ has been modeled extensively using fundamental chemistry knowledge. For example, there are classical mechanics-based force fields, such as the widely-used Merck Molecular Force Field (MMFF) (Halgren, 1996), and quantum mechanical-based approaches, such as Density Functional theory (DFT).

Generating accurate samples from this distribution is essential for computing accurate molecular property expectations. However, sampling from $p(x)$ is very challenging because it is a high-dimensional and multi-modal distribution; generating samples from high-dimensional distributions is a classic problem in computational statistics (Liu, 2008).

The complexity in the Boltzmann distribution is primarily due to the large number of degrees of freedom that a given molecule possesses. These degrees of freedom are typically defined in terms of three types of so-called internal coordinates: bond lengths, which are distances between pairs of bonded atoms; bond angles, which are angles between bonds connected to a common atom; and torsions, which are dihedral angles formed by four consecutively bonded atoms. The number of degrees of freedom therefore increases substantially as a function of molecule size.

There are a variety of classical approaches to conformation generation, including stochastic methods and systematic methods (Hawkins, 2017). Some stochastic methods, such as Markov Chain Monte Carlo and Molecular Dynamics, attempt to generate samples from the Boltzmann distribution (which we refer to below as "Boltzmann samples") but can suffer from long runtimes on the order of minutes or even

[1]Department of Computer Science, University of Chicago, Chicago, USA. Correspondence to: Kirk Swanson <swansonk1@uchicago.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

hours for large molecules (Tuckerman, 2010).

Other stochastic methods, such as Monte Carlo Multiple Minimum and Low Mode Search, generate low energy (i.e., high probability) conformations and have shorter runtimes (Chang et al., 1989; Freyberg & Braun, 1991; Wilson et al., 1991; Kolossvary & Guida, 1996; Parish et al., 2002; Supady et al., 2015; Schrödinger, 2016). However, finding *modes* is different than generating true *samples* from the underlying distribution. Generating modes can be useful, and it is possible to utilize strategies such as importance sampling to compute expectations with them. However, this is not ideal for computing expectations with respect to the Boltzmann distribution, because there are lots of inherent symmetries and flexibility that can make mere mode identification insufficient.

Distance geometry methods, such as RDKit's ETKDG algorithm (Riniker & Landrum, 2015), use a stochastic approach in conjunction with knowledge-based rules to produce a diverse set of geometries, but these conformations are not Boltzmann samples. Systematic methods, such as Confab (O'Boyle et al., 2011), aim to generate conformations via brute force search or using knowledge-based libraries, but this becomes intractable for large molecules and these methods also do not produce Boltzmann samples (Kothiwale et al., 2015; Cole et al., 2018).

A number of machine learning approaches have been introduced which aim to improve the speed and accuracy of conformation generation. Most of these approaches, including DL4Chem (Mansimov et al., 2019), CGCF (Xu et al., 2021), GeoMol (Ganea et al., 2021), and RMCF (Wang et al., 2022), have focused on generating low energy conformations (Shi et al., 2021; Luo et al., 2021; Xu et al., 2022). Similar to some of the classical methods, these methods do not generate Boltzmann samples. GraphDG was designed to generate such samples when combined with an importance sampling scheme, but it was only trained and evaluated on a limited set of 197 distinct molecular graphs from a single molecular formula (Simm & Hernández-Lobato, 2020). Noé *et al.* devised so-called Boltzmann generators, which utilize normalizing flows to generate Boltzmann samples (Noé et al., 2019). These have shown promise on proteins as well as small organic molecules (Köhler et al., 2021), but a separate model needs to be trained for every molecule, because the normalizing flows operate on intrinsic coordinates specific to a given molecule. This limits their utility for high-throughput tasks (Jing et al., 2022).

A Torsional Diffusion method was recently introduced by Jing *et al.* which exclusively models rotatable bond torsion angles (Jing et al., 2022). It is well-understood that the primary source of conformational variability arises from torsional rotations about so-called rotatable bonds (Chen & Foloppe, 2011; Axelrod & Gomez-Bombarelli, 2022). Rotatable bonds are canonically defined as single, non-ring bonds which are attached to atoms that are non-terminal and not triply bonded.[1] Jing *et al.* use a slightly more expansive definition of rotatable bond as any bond which, if severed, would produce two connected components in the molecular graph. The authors trained a Torsional Diffusion model to produce low energy conformations, and it showed state-of-the-art performance on standard datasets. Because the model provides exact likelihoods, the authors were also able to train a Torsional Diffusion model to generate independent samples from the marginal Boltzmann distribution of rotatable bond torsion angles for a variety of molecules. However, it was only trained and evaluated on molecules with between 3 and 7 rotatable bonds (using the authors' definition), which excludes large swaths of chemical space; ibuprofen, for example, has 8 rotatable bonds. Torsional Diffusion is also more computationally expensive than other machine learning models.

In this work, we present VonMisesNet, a new graph neural network that captures conformational variability via a variational approximation of rotatable bond torsion angles as a mixture of von Mises distributions.[2] Similar to Torsional Diffusion, we focus on modeling the rotatable bonds, although we use the canonical rotatable bond definition because it excludes some bonds, such as double or triple bonds, that may not be freely rotatable. VonMisesNet places no restriction on the number of rotatable bonds that it can process, and therefore it can be used to generate conformations for arbitrary molecular graphs. It is also the first machine learning method that specifically accounts for chirality inversion, a phenomenon that can strongly influence the local geometry about atoms with three neighbors and one lone pair of electrons, which is often the case with nitrogen atoms. We show that our method is not only more accurate than other methods in terms of its ability to generate samples from the marginal Boltzmann distribution of rotatable bond torsion angles, but that it is also orders of magnitude faster than diffusion-based methods.

2. Methods

2.1. Ground Truth Conformation Generation

To train a machine learning model to generate Boltzmann samples, we require datasets that contain Boltzmann samples for a variety of molecules. However, most widely-used benchmark datasets consist of low energy conformations. For example, the popular datasets GEOM-QM9 (Ramakrishnan et al., 2014) and GEOM-DRUGS (Axelrod & Gomez-

[1] This definition is based on the canonical rotatable bond SMARTS string from RDKit: `[!$(#)&!D1]-!@[!$(#)&!D1]`.

[2] Code is available at https://github.com/thejonaslab/vonmises-icml-2023.

Bombarelli, 2022) consist of low energy conformations generated from DFT calculations. Simm and Hernandez-Lobato introduced the CONF17 benchmark (2020), which consists of conformations generated via *ab initio* Molecular Dynamics simulations. Although these conformations are Boltzmann samples, the dataset only consists of 197 unique molecular graphs with the formula $C_7H_{10}O_2$.

Therefore, to train and evaluate on Boltzmann samples that represent a much larger and more diverse set of molecules, we construct our own ground truth dataset. We combine Parallel Tempering and Hamiltonian Monte Carlo, techniques that are both well-known to efficiently sample complex distributions (Liu, 2008; Tuckerman, 2010). We refer to this method as PT-HMC (see Appendix §A for details).

In our PT-HMC simulations, we use MMFF to give the molecular potential energy, $E(x)$. The Torsional Diffusion model that was trained to generate Boltzmann samples for rotatable bonds also utilized MMFF. We use the MMFF implementation in RDKit to compute potential energies and gradients (Tosco et al., 2014). Although quantum mechanical approaches would yield more physically accurate geometries, classical force fields allow for more computationally efficient simulations while still providing good approximations to the true Boltzmann distribution. This lets us show proof of concept for our method; future work will investigate the use of quantum mechanical approaches.

2.2. Datasets

We used PT-HMC to generate conformations for two datasets of molecules: NMRShiftDB and GDB-17 (see details in Appendix §B). We use 32,171 molecules from NMRShiftDB and 134,228 molecules from GDB-17 that have at most 64 atoms and elements in the set {H, C, O, N, F, S, P, Cl}. We split each of these datasets into training (NMRShiftDB-train and GDB-17-train) and test (NMRShiftDB-test and GDB-17-test) datasets by computing a hash of the Morgan fingerprint with a radius of 4 and 2,048 bits for each molecule. If the last digit of this hash is a 0 or a 1, the molecule is in the test set, otherwise it is in the train set. This produces an 80/20 train/test split, and it allows us to consistently determine whether a given molecule is in the train set or the test set. Each molecule has approximately 560 conformations.

2.3. VonMisesNet

In this section, we describe VonMisesNet and how it is used to generate conformations. In subsections 2.3.1, 2.3.2, and 2.3.3, we describe the geometric components that VonMisesNet predicts and explain our modeling choices. In subsection 2.3.4, we describe the VonMisesNet architecture and training procedure. In subsection 2.3.5, we explain how the predictions from VonMisesNet can be used to rapidly generate accurate molecular conformations.

2.3.1. Rotatable Bond Modeling

Our primary goal is to model the marginal Boltzmann distribution of rotatable bond torsion angles in a given molecule. To do so, we need a canonical way of defining these angles, which are dihedral angles formed by four consecutively bonded atoms. The central two atoms are fixed as the begin and end atoms of the bond itself, but there can be multiple choices for the other two atoms. We use a breadth-first search approach based on the Cahn-Ingold-Prelog (CIP) (Cahn et al., 1966) rules to determine which four atoms, including the two central atoms, should be chosen (see Appendix §C for details).

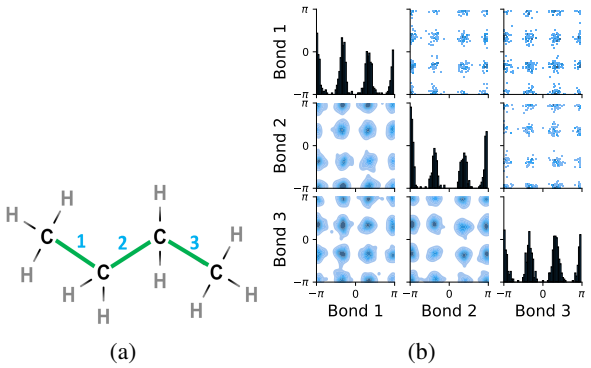


Figure 1. Rotatable bonds example. **(a)** Molecular graph of butane, with the three rotatable bonds highlighted. **(b)** Marginal and pairwise joint distributions of the three rotatable bond torsion angles in butane from 560 conformations generated via PT-HMC. The diagonal plots show the marginal distributions. Each of these distributions has three distinct modes (note that the axes represent a periodic angle variable ranging from $-\pi$ to π). The plots above and below the diagonal show scatter and kernel density estimates of the pairwise joint distributions, respectively. The pairwise joint distributions present clearly as product distributions, which indicates that the distributions are approximately independent.

If θ_i is the angle of the i^{th} rotatable bond and there are N rotatable bonds, then our target distribution is given by $p(\theta_1, ..., \theta_N)$. We make two primary assumptions about this distribution: independence and multi-modality.

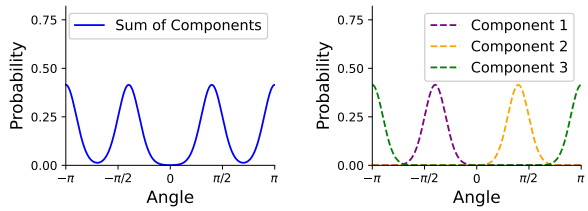
Independence. The first assumption is that the individual rotatable bond torsion angle distributions are approximately independent. Under this assumption, we express the target distribution as a product:

$$p(\theta_1, ..., \theta_N) \approx \prod_{i=1}^{N} p(\theta_i) \tag{2}$$

We find that this formulation facilitates straightforward modeling and rapid generation of conformations, but in future

work we hope to relax this assumption and to explicitly model long-range interactions.

Butane is an example of a molecule that supports this assumption. In Figure 1, the pairwise joint distributions of the rotatable bond torsion angles clearly present as product distributions. In Appendix §D, we show that these distributions are approximately rank-1, which indicates independence. We also show that on average, for a larger set of molecules, pairs of rotatable bonds have a joint distribution that is approximately rank-1, with the approximation improving when the bonds are farther apart within a molecule.



(a) (b)

Figure 2. von Mises distribution example. **(a)** The probability density of a mixture of three von Mises distributions with means $\mu_1 = -1.25$, $\mu_2 = 1.25$ and $\mu_3 = \pi$. The concentrations κ_i are all equal to 10, and the weights w_i are all equal to 1/3. **(b)** The probability densities of the individual component distributions.

Multi-modality. The second assumption is that most individual rotatable bond torsion angle distributions are multi-modal, typically with up to four distinct modes. As shown in Figure 1, butane also exemplifies this. In Appendix §E, we show that in a larger set of molecules, the majority of rotatable bond torsion angle distributions have fewer than four distinct modes.

With these assumptions, we model a given rotatable bond torsion angle distribution as a mixture of von Mises distributions. We use a von Mises distribution because it is a continuous probability distribution on the circle with support in $[-\pi, \pi]$, which is the periodic range of rotatable bond torsion angles. We use a mixture of von Mises distributions to capture the multi-modal nature of rotatable bond torsion angles. The weighted sum of N von Mises distributions is given by

$$p(\theta) = \sum_{i=1}^{N} w_i \frac{e^{\kappa_i \cos(\theta - \mu_i)}}{2\pi \mathrm{I}_0(\kappa_i)}, \tag{3}$$

where θ is the angle, I_0 is the modified Bessel function of order 0, w_i is the weight, μ_i is the mean, and κ_i is the concentration of the i^{th} distribution.[3] Figure 2 shows an example. In our experiments we use $N = 4$ as most of the distributions have up to four distinct modes. VonMisesNet predicts the parameters of this weighted sum for each rotatable bond using the molecular graph as the input. Hence, we capture conformational variability via a variational approximation of rotatable bond torsion angles as a mixture of von Mises distributions.

[3]The concentration is analagous to the inverse variance.

2.3.2. Chirality Inversion Modeling

Some atoms can undergo a specific type of transformation that influences the local geometry about a rotatable bond. Atoms that have three bonded neighbors and one lone pair of electrons can exhibit chirality inversion. (From here on we refer to such atoms as "chirality inversion atoms.") This means that the atom moves through the plane formed by its three neighbors, which causes an oscillation between R and S chirality (see Appendix §G for details). This can often occur with nitrogen atoms, where the thermodynamic barrier for this inversion, $\sim$ 25 kJ/mol, is low enough to allow rapid inversion at room temperature, leading to a racemic mixture of R and S configurations (Kennepohl, 2022). An example is shown in Figure 3.

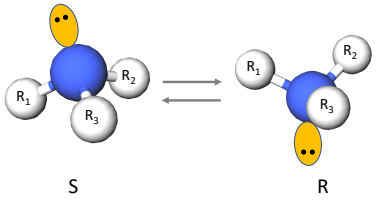


Figure 3. Chirality inversion example. A molecular fragment with a chirality inversion atom (blue) connected to three bonded neighbors (white) along with a lone pair (yellow). The chirality inversion atom moves through the plane formed by the three neighbors. Assuming that the neighbors are numbered in order of decreasing CIP priority and that the lone pair has the lowest priority, the configuration on the left has S chirality and the configuration on the right has R chirality; one cannot simply rotate the one on the left to achieve the one on the right.

If a chirality inversion atom is an endpoint of a rotatable bond, then the inversion can influence the rotatable bond torsion angle distribution. 31% of molecules in our NMRShiftDB data and 60% of molecules in our GDB-17 data have at least one rotatable bond connected to a chirality inversion atom.

We model chirality inversion as follows. Let θ be a rotatable bond torsion angle where one of the endpoint atoms, A, is a chirality inversion atom.[4] Let $p(\theta|R)$ be the angle distribution when A has R chirality, $p(\theta|S)$ be the angle distribution

[4]There are some cases where both of the atom endpoints are chirality inversion atoms. There are only 203 molecules with such rotatable bonds in our NMRShiftDB data (0.6%) and none in our GDB-17 data, so we do not explicitly handle this special case. As described below, when making predictions for such cases, we condition on the chirality of just one of the endpoint atoms.

when A has S chirality, and $p(R)$ be the probability that A has R chirality. Then, we can express the rotatable bond torsion angle distribution as

$$p(\theta) = p(R) \cdot p(\theta|R) + (1 - p(R)) \cdot p(\theta|S). \quad (4)$$

We model $p(\theta|R)$ and $p(\theta|S)$, each, as mixtures of von Mises distributions, as described above. For each rotatable bond that has a chirality inversion endpoint atom, VonMisesNet predicts the von Mises mixture parameters for both $p(\theta|R)$ and $p(\theta|S)$. It additionally predicts $p(R)$ for each chirality inversion atom in the molecule.

2.3.3. Bond Length and Bond Angle Modeling

Although rotatable bonds are responsible for most of the conformational variability in a molecule, its geometry is also determined by bond lengths and bond angles. These are typically unimodal distributions with minor variance, which obviates the need to predict full distributions for them. Therefore, VonMisesNet predicts averages for bond lengths and bond angles.

2.3.4. Architecture and Training

VonMisesNet is a graph neural network that takes a molecular graph with an initial 3D structure as input and predicts the following: the probability of R chirality for each chirality inversion atom, average length for all bonds, average angle for all bond angles, and parameters for a mixture of four von Mises distributions for each rotatable bond (including two sets of parameters for any rotatable bond that is connected to a chirality inversion atom, conditioned on R and S). These predictions can be used to rapidly generate conformations. We use a multi-partite graph representation for molecules, which removes the need to update edge or global features, allowing a simple message passing scheme to be used for nodel-level predictions of all quantities of interest. In Figure 4, we illustrate the multi-partite graph representation and summarize the neural network architecture that we use to produce node-level predictions (see Appendix §F for details on our node-level featurization and Appendix §J for further details on the neural network architecture). During training, we minimize the negative log likelihood of the ground truth angle samples for a given rotatable bond under a mixture of four von Mises distributions defined by the predicted parameters. For all of the other predictions, we minimize the mean squared error.

2.3.5. Conformation Generation

To generate conformations, we create an initial geometry from ETKDG and run inference with VonMisesNet on the molecular graph. Using RDKit, we set all of the non-ring bonds in the initial geometry, as well as the bond angles containing at least one rotatable bond, to the respective predicted averages from VonMisesNet. Then, we generate a specified number of conformations using the following steps. For each chirality inversion atom, we randomly set the atom to have R or S chirality based on the predicted probability (see Appendix §G for details). For each rotatable bond, we sample a value from its predicted von Mises mixture distribution and set the rotatable bond torsion angle in the initial geometry to that value. For rotatable bonds that are connected to a chirality inversion atom, we use the von Mises mixture distribution that is conditioned on the chirality of that atom. When both endpoints of the bond are chirality inversion atoms, we condition on the chirality of the "begin atom" in RDKit. At this point, a full conformation has been generated. This process is repeated to generate the specified number of conformations.

There is an optional filtering step in this process. After a conformation is generated, we check if there are any two atoms separated by more than five bonds that are closer than the average of their van der Waals radii. If so, we discard the conformation and re-generate, up to a maximum number of total retries. This is inspired by the ETKDG algorithm, which uses the sum of the van der Waals radii of two atoms more than five bond lengths apart as a constraint in the distance geometry process.[5] These constraints are useful because it is unphysical for two atoms to overlap spatially. We find that this filtering approach is a crude but effective way of mitigating the fact that VonMisesNet does not explicitly model long-range interactions.

3. Results

In this section, we evaluate the speed and accuracy of VonMisesNet. In subsection 3.1 we discuss baselines for comparison, and in subsection 3.2 we compare speed via runtime per conformer. For the problem of generating low energy conformations, accuracy evaluations typically focus on a set of reference conformations and use RMSD to measure the fraction that are recovered by a given method. Because our goal is to sample from a distribution rather than to generate low energy conformations, we use accuracy metrics that compare distributions rather than individual conformations. Specifically, we compare the generated distributions of rotatable bond torsion angles (subsection 3.3) and the distributions of pairwise atomic distances (subsection 3.4) to the respective ground truth Boltzmann distributions.

3.1. Baselines

We use three baselines: ETKDG, GeoMol, and Torsional Diffusion. When ETKDG generates conformations, it oc-

[5]We use a looser constraint as we find that it increases the ratio of preserved conformations while also improving evaluation metrics.

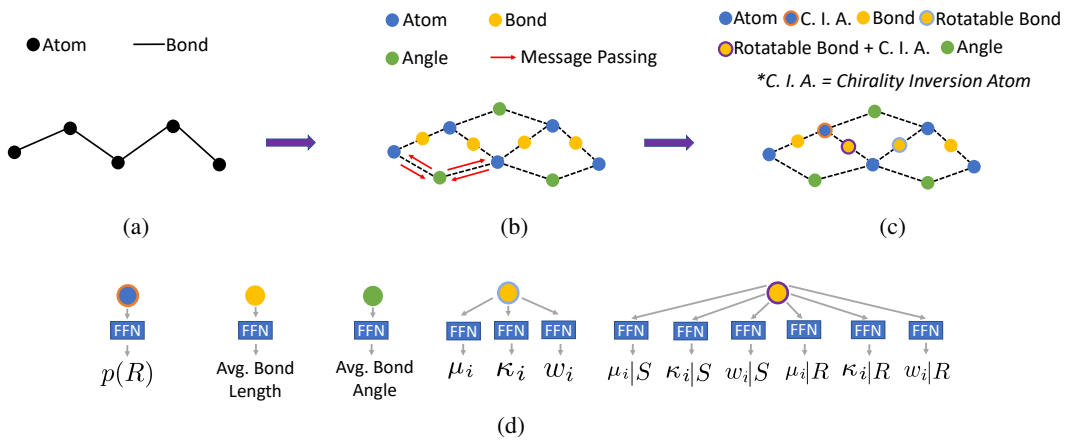


Figure 4. VonMisesNet architecture. **(a)** A fragment of a molecule represented in typical fashion with atoms corresponding to nodes and bonds corresponding to edges. **(b)** We transform this graph into a multi-partite representation that has atom nodes, bond nodes, and angle nodes. The atoms remain fixed as nodes, shown in blue, but the bond edges become two edges connected to a bond node, shown in yellow. In addition, we add angle nodes, shown in green, which are connected by edges to the endpoints of a set of three consecutive atom nodes that define a bond angle. We perform message passing on this input graph, which results in node-level vectors. For each node type, shown in **(c)**, we generate the relevant predictions using Feed Forward Networks (FFN), shown in **(d)**. $p(R)$ is the probability of R chirality, and μ_i, k_i, and w_i are the von Mises parameters.

casionally swaps the indices of otherwise indistinguishable atoms, which can affect the comparison metrics we use. We made a small modification to enforce consistent atom indexing, which we call ETKDG-Clean (see Appendix §I for details). For Torsional Diffusion, we use the Boltzmann generator model that was trained at 300 Kelvin, using the default 20 de-noising steps, a target sample temperature of 293 Kelvin, and the GEOM-DRUGS featurization configuration. Although GeoMol is designed to generate low energy conformations, we use it as a baseline to examine whether such a model can be easily repurposed to generate Boltzmann samples. Because GeoMol was originally trained on DFT geometries, we re-trained it on our ground truth data using the default hyperparameters and GEOM-DRUGS featurization configuration so that it generates geometries that are closer to our MMFF-based ground truth. When evaluating on molecules from NMRShiftDB-test, we use the VonMisesNet and GeoMol models that were trained on NMRShiftDB-train, and when evaluating on GDB-17-test, we use the corresponding models that were trained on GDB-17-train. Below, VonMisesNet-Filtered means that we use the optional filtering step when generating conformations. We generate 560 conformations per molecule for each method.

3.2. Runtime

Runtime is an important metric because conformation generation is often an intermediate step in high-throughput molecular property prediction tasks. In Table 1, we compare the average time it takes to generate a single conformation. VonMisesNet is the fastest method and Torsional Diffusion is the slowest. Compared to Torsional Diffusion, VonMisesNet is about 47 times faster on a GPU and about 220 times faster on a CPU.

Table 1. Runtime per conformer in milliseconds averaged across 58 molecules from NMRShiftDB-test, along with standard error measurements. We selected 100 random molecules, 58 of which satisfied the Torsional Diffusion rotatable bond constraints. 100 conformers were generated for each molecule on a 64-core machine that has a single NVIDIA GeForce RTX 2080 Ti GPU. We do not include the start up times for any of the methods (e.g., loading neural network weights).

METHOD	CPU (MS)	GPU (MS)
ETKDG-CLEAN	12.3 ± 1.5	NA
GEOMOL	3.6 ± 0.2	NA
TORSIONAL DIFFUSION	682.7 ± 46.9	140.0 ± 5.3
VONMISESNET	**3.1** ± **0.2**	**3.0** ± **0.2**
VONMISESNET-FILTERED	4.7 ± 0.7	4.9 ± 0.8

3.3. Rotatable Bond Torsion Angle Distributions

To evaluate the accuracy of generated rotatable bond torsion angle distributions, we measure the KL divergence and Earth Mover's Distance (EMD) relative to the PT-HMC ground truth. These metrics show how well a model is capturing the distributions for individual rotatable bonds. Among 1000 random molecules from NMRShiftDB-test, we select 538 that satisfy the Torsional Diffusion rotatable bond constraints, and among 1000 random molecules from

GDB-17-test, we select 610 that satisfy the constraints.[6] As shown in Table 2, VonMisesNet and VonMisesNet-Filtered outperform the other methods and have the lowest average KL and EMD values. In Figure 5, we examine a specific example where chirality inversion has a nontrivial effect on the distribution, which is captured by VonMisesNet. In Figure 6, we compare VonMisesNet and PT-HMC across several example rotatable bonds. Additional evaluations on the full sets of 1000 random molecules, without the Torsional Diffusion constraints, are shown in Appendix §H.

Table 2. Rotatable bond torsion angle distributions evaluation. Average KL divergence and EMD of rotatable bond torsion angle distributions, per molecule, relative to PT-HMC ground truth in 538 molecules from NMRSfhiftDB-test and 610 molecules from GDB-17-test, respectively. Standard error is in parentheses. The KL is measured with 32 bins. EC, GM, TD, VMN, and VMN-F stand for ETKDG-Clean, GeoMol, Torsional Diffusion, VonMisesNet, and VonMisesNet-Filtered, respectively. NMR and GDB stand for NMRShiftDB-test, and GDB-17-test, respectively.

	KL Divergence				
	EC	GM	TD	VMN	VMN-F
NMR	2.80 (0.09)	2.62 (0.08)	1.14 (0.05)	0.82 (0.04)	**0.76 (0.04)**
GDB	3.97 (0.09)	2.70 (0.07)	1.71 (0.04)	1.36 (0.04)	**1.31 (0.04)**
	Earth Mover's Distance (EMD)				
	EC	GM	TD	VMN	VMN-F
NMR	0.75 (0.03)	0.92 (0.04)	0.76 (0.03)	0.47 (0.03)	**0.46 (0.02)**
GDB	1.13 (0.03)	1.69 (0.07)	1.02 (0.03)	**0.93 (0.03)**	**0.93 (0.03)**

3.4. Pairwise Distance Distributions

To evaluate the accuracy of pairwise distance distributions, we measure the EMD of these distributions as well as the mean absolute error (MAE) of the expected distance relative to the PT-HMC ground truth.[7] When the shortest path between two atoms contains n atoms, we say that the distance is a 1-n distance. Comparing these 1-n distance distributions gives us a way to evaluate the accuracy of generated geometries as a function of the graph-distance between pairs of atoms. We use the same sets of molecules as in §3.3. In the first row in Figure 7, we evaluate 1-n distances for which every intermediate bond along the shortest path is rotatable, which lets us focus on how well models are capturing the marginal Boltzmann distribution of rotatable bond torsion angles.[8] VonMisesNet, VonMisesNet-Filtered, and Torsional Diffusion outperform the other methods. VonMisesNet shows relatively stronger performance for $n < 6$ and weaker performance in some cases for $n = 6$, and VonMisesNet-Filtered performs best overall. In the second row, we allow for any intermediate bond along the shortest path except for those that belong to a non-aromatic ring. There is a regression in the relative performance of GeoMol and Torsional Diffusion, and VonMisesNet and VonMisesNet-Filtered outperform all other methods on nearly all metrics. In the third row, we remove the non-aromatic restriction and consider all bonds along the shortest path, and ETKDG-Clean performs better overall. Future work will handle non-aromatic rings, which are an especially complex case for which, to the best of our knowledge, no full machine learning solution exists. Evaluations with $n > 6$ and without the Torsional Diffusion constraints are shown in Appendix §H.

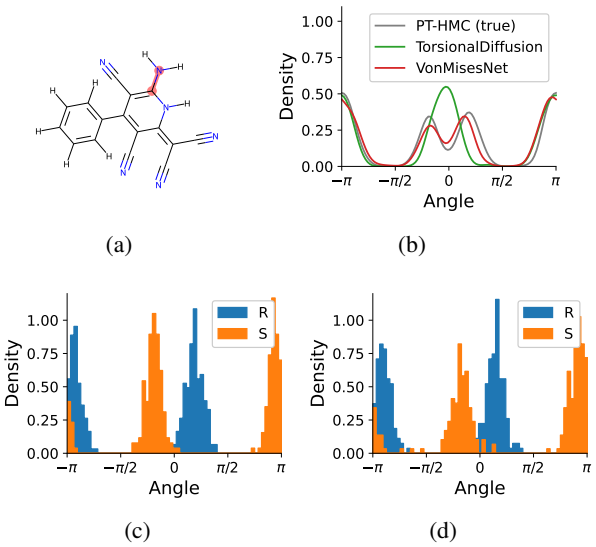


Figure 5. Predicting chirality inversion example. **(a)** A molecule from NMRShiftDB-test, where the highlighted rotatable bond is between a carbon atom and a chirality inversion nitrogen atom. The nitrogen atom has R chirality in 54.2% of the PT-HMC conformations, 47.8% of the VonMisesNet conformations, and 100% of the Torsional Diffusion conformations. **(b)** Kernel density estimates of the rotatable bond torsion angle distribution from PT-HMC, VonMisesNet, and Torsional Diffusion. **(c)** and **(d)** show the distributions of this angle conditioned on the chirality of the nitrogen atom for PT-HMC and VonMisesNet, respectively. See Figure 21 in the Appendix for a plot that includes ETKDG-Clean and GeoMol as well.

4. Discussion

In this work, we presented VonMisesNet, a graph neural network that models conformational variability with a variational approximation of rotatable bond torsion angles as a mixture of von Mises distributions. Conformations generated with VonMisesNet have more accurate rotatable bond torsion angle distributions with respect to the Boltzmann

[6] We do not use the entire test sets due to the large amount of time it would take to run inference for all methods.

[7] We do not use KL because there are no consistent lower and upper bounds on the distances, and therefore we cannot easily assign bins for KL.

[8] By intermediate bonds, we mean all bonds along the shortest path except for the first and the last.

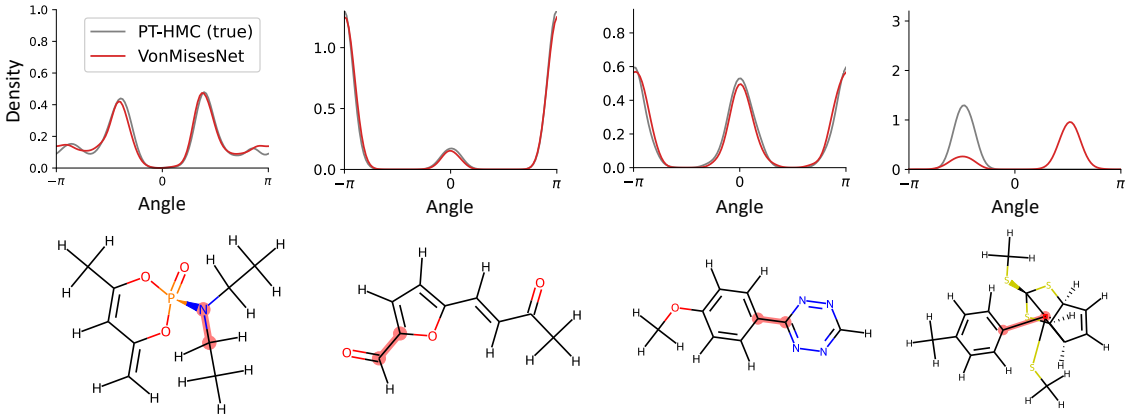


Figure 6. VonMisesNet prediction examples. We compare kernel density estimates of the torsion angle distributions of several rotatable bonds from PT-HMC and VonMisesNet. These molecules are taken from NMRShiftDB-test. The KL divergence of the VonMisesNet distributions relative to PT-HMC are, from left to right, 0.11, 0.024, 0.25, and 3.19. VonMisesNet performs well in the first three and poorly in the fourth. In the fourth example, the PT-HMC distribution is an odd function despite the presence of a methylbenzene group, which is symmetric under flipping of the aromatic ring. We find that the energy barrier is approximately 800 kcal/mol based on the MMFF implementation in RDKit, suggesting that such 180 flips may never occur in nature.

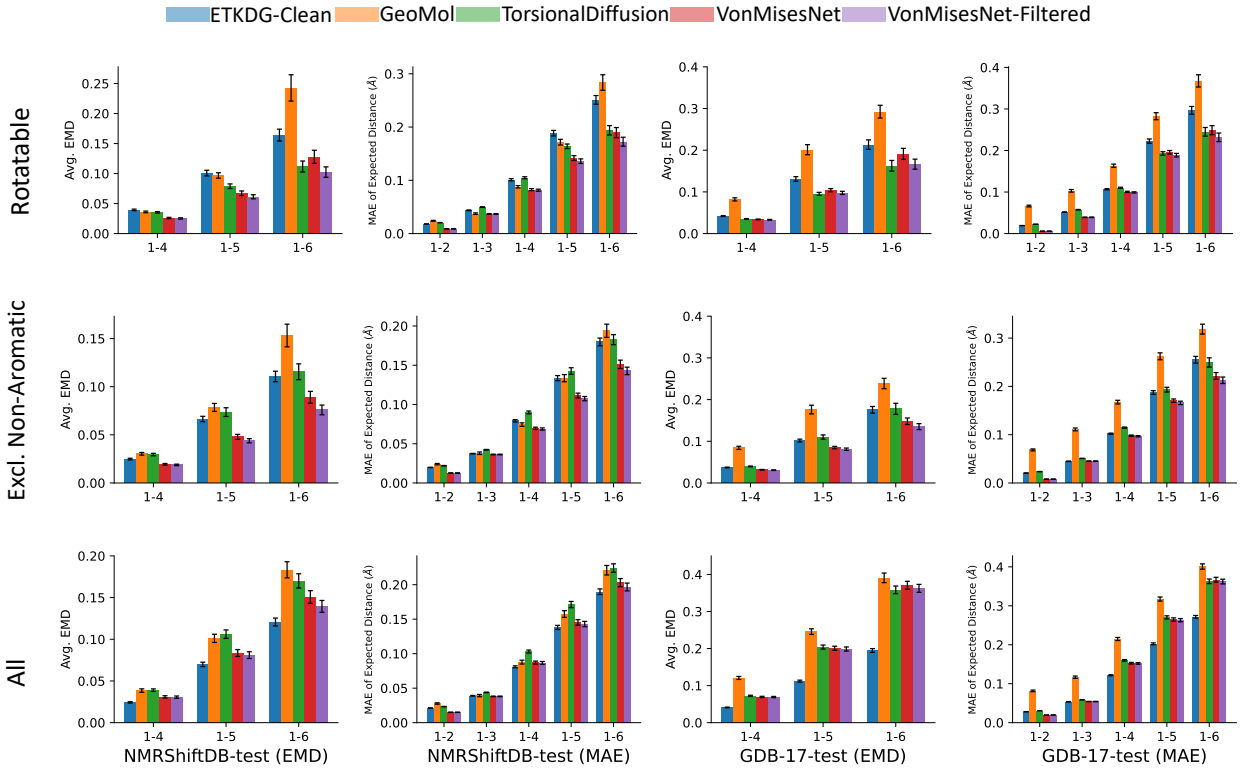


Figure 7. Pairwise distance distributions evaluation. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 538 molecules from NMRShiftDB-test and 610 molecules from GDB-17-test. We evaluate 1-4, 1-5, and 1-6 distances for which every intermediate bond along the shortest path is rotatable in row 1, for which every intermediate bond along the shortest path is not part of a non-aromatic ring in row 2, and with no restrictions in row 3. In columns 1 and 2 we compare the average EMD, per molecule, and the MAE of expected distance, per molecule, respectively, for the molecules from NMRShiftDB-test. Columns 3 and 4 show the same for molecules from GDB-17-test. For the expected distance evaluations, in row 1 we additionally include 1-2 distances that are not part of a ring and 1-3 distances for which at least one of the bonds is rotatable. In row 2, we include 1-2 distances that are not part of a non-aromatic ring and 1-3 distances for which neither of the bonds is in a non-aromatic ring. In row 3, there are no 1-2 or 1-3 restrictions. We exclude 1-2 and 1-3 distances from the EMD evaluations as these distributions are typically unimodal with small variance.

distribution than the other methods, and they also tend to have the most accurate 1-n pairwise distance distributions, especially for $n < 7$ and when excluding non-aromatic rings. The performance of Torsional Diffusion tends to improve relative to VonMisesNet as a function of n, but VonMisesNet is orders of magnitude faster. To the best of our knowledge, it is also the only machine learning model that takes chirality inversion into account.

There are several avenues for future work. First, we focused exclusively on rotatable bonds and have not yet modeled non-aromatic rings, which are often flexible and can contribute to conformational variability. Second, we showed proof of concept with MMFF-based energies, but training on ground truth data from quantum mechanical calculations would allow for generating conformations with more accurate geometries. Finally, explicitly modeling joint rotatable bond torsion angle distributions would likely yield more accurate long-range interactions. Torsional Diffusion achieves this with a diffusion model that operates on the hypertorus defined by the torsion angles. Some of the low energy conformation generation methods also model joint probabilities. For example, GeoMol jointly predicts all torsion angles with graph neural networks, and RMCF models the joint distribution of molecular fragments and dihedral angles via a Markov random field. However, these methods are not suitable for Boltzmann sampling.

We believe that the variational inference framework introduced here could be extended to joint distributions, and therefore long-range interactions, while maintaining computational efficiency. For example, future ideas include modeling the joint distribution for all pairs of rotatable bond torsion angles, yielding a Markov random field which is still fast to sample from, in contrast to diffusion approaches that require multiple slow diffusion steps (multiple neural network forward passes) for each conformation.

Acknowledgements

The authors thank all of the members of the Jonas Lab (University of Chicago), as well as Kyle Swanson and Melody Huang, for discussions and advice. This material is based upon work supported by the National Science Foundation under Award 2231634. Computing research and services were partly provided by the OSG Consortium which is supported by the National Science Foundation awards 2030508 and 1836650 and by Texas Advanced Computing Center (TACC) at the University of Texas, Austin via award CHE21008.

References

Axelrod, S. and Gomez-Bombarelli, R. GEOM, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(185), 2022.

Cahn, R. S., Ingold, C. K., and Prelog, V. Specification of Molecular Chirality. *Angewandte Chemie International Edition*, 5(4):385–415, 1966.

Chang, G., Guida, W. C., and Still, W. C. An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *Journal of the American Chemical Society*, 111: 4379–4386, 1989.

Chen, I. J. and Foloppe, N. Is Conformational Sampling of Drug-like Molecules a Solved Problem? *Drug Development Research*, 72:85–94, 2011.

Cole, J. C., Korb, O., McCabe, P., Read, M. G., and Taylor, R. Knowledge-Based Conformer Generation Using the Cambridge Structural Database. *Journal of Chemical Information and Modeling*, 58:615–629, 2018.

Freyberg, B. V. and Braun, W. Efficient Search for All Low Energy Conformations of Polypeptides by Monte Carlo Methods. *Journal of Chemical Information and Modeling*, 12(9):1065–1076, 1991.

Ganea, O.-E., Pattanaik, L., Coley, C. W., Barzilay, R., Jensen, K., Green, W., and Jaakkola, T. S. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=af_hng9tuNj.

Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94y. *Journal of Computational Chemistry*, 17(5 and 6):490–519, 1996.

Hawkins, P. C. D. Conformation Generation: The State of the Art. *Journal of Chemical Information and Modeling*, 57:1747–1756, 2017.

Jing, B., Corso, G., Chang, J., Barzilay, R., and Jaakkola, T. S. Torsional diffusion for molecular conformer generation. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=w6fj2r62r_H.

Kennepohl, D. Stereochemistry at Tetrahedral Centers. LibreTexts, 2022. URL https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Organic_Chemistry_(LibreTexts)/05%3A_Stereochemistry_at_Tetrahedral_Centers.

Köhler, J., Krämer, A., and Noe, F. Smooth normalizing flows. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=yxsak5ND2pA.

Kolossvary, I. and Guida, W. C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *Journal of the American Chemical Society*, 118:5011–5019, 1996.

Kothiwale, S., Mendelhall, J. L., and Meiler, J. BCL::Conf: small molecule conformational sampling using a knowledge based rotamer library. *Journal of Cheminformatics*, 7(47), 2015.

Kuhn, E. J. . S. Rapid prediction of NMR spectral properties with quantified uncertainty. *Journal of Cheminformatics*, 11(50), 2019.

Liu, J. S. *Monte Carlo Strategies in Scientific Computing*. Springer, 2008.

Luo, S., Shi, C., Xu, M., and Tang, J. Predicting molecular conformation via dynamic graph score matching. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=hMY6nm9lld.

Mansimov, E., Mahmood, O., Kang, S., and Cho, K. Molecular Geometry Prediction using a Deep Generative Graph Neural Network. *Scientific Reports*, 9(20381), 2019.

Noé, F., Olsson, S., Köler, J., and Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), 2019.

O'Boyle, N. M., Vandermeersch, T., Flynn, C. J., Maguire, A. R., and Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *Journal of Cheminformatics*, 3(8), 2011.

Parish, C., Lombardi, R., Sinclair, K., Smith, E., Goldberg, A., Rappley, M., and Dure, M. A comparison of the Low Mode and Monte Carlo conformational search methods. *Journal of Molecular Graphics and Modelling*, 21:129–150, 2002.

Ramakrishnan, R., Dral, P. O., Rupp, M., and Lilienfeld., O. A. V. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Riniker, S. and Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *Journal of Chemical Information and Modeling*, 55(12):2562–2574, 2015.

Ruddigkeit, L., van Deursen, R., Blum, L. C., and Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *Journal of Chemical Information and Modeling*, 52(11): 2864–2875, 2012.

Schrödinger. Macromodel, 2016.

Shi, C., Luo, S., Xu, M., and Tang, J. Learning gradient fields for molecular conformation generation. In *Proceedings of the 38 th International Conference on Machine Learning*, 2021.

Simm, G. N. C. and Hernández-Lobato, J. M. A generative model for molecular distance geometry, 2020. URL https://openreview.net/forum?id=S1x1IkHtPr.

Supady, A., Blum, V., and Baldauf, C. First-Principles Molecular Structure Search with a Genetic Algorithm. *Journal of Chemical Information and Modeling*, 55:2338–2348, 2015.

Tosco, P., Stiefl, N., and Landrum, G. Bringing the MMFF force field to the RDKit: implementation and validation. *Journal of Cheminformatics*, 6(37), 2014.

Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*. John Wiley and Sons, 2010.

Wang, L., Zhou, Y., Wang, Y., Zheng, X., Huang, X., and Zhou, H. Regularized molecular conformation fields. In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K. (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=7XCFxnG8nGS.

Wilson, S. R., Cui, W., Moskowitz, J. W., and Schmidt, K. E. Applications of Simulated Annealing to the Conformational Analysis of Flexible Molecules. *Journal of Computational Chemistry*, 12(3):342–349, 1991.

Xu, M., Luo, S., Bengio, Y., Peng, J., and Tang, J. Learning neural generative dynamics for molecular conformation generation. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=pAbm1qfheGk.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

A. Simulation Details for Ground Truth Conformation Generation

To generate ground truth molecular conformations, we combine Parallel Tempering and Hamiltonian Monte Carlo (PT-HMC). In the context of molecular systems, Hamiltonian Monte Carlo is a Monte Carlo simulation where at each step, a short Molecular Dynamics simulation is used to propose a molecular geometry. Parallel Tempering refers to the technique of running multiple Monte Carlo simulations simultaneously at different temperatures and periodically proposing swaps between the states of adjacent temperatures. For the PT-HMC simulations, we used 70,000 total steps, sampling every 100 steps. We used eight different temperatures for parallel tempering: 293, 400, 500, 600, 700, 800, 900, and 1000 Kelvin. 293 Kelvin served as the target base temperature. In other words, the generated conformations are sampled from the Boltzmann distribution with $T = 293$ Kelvin. We used a Parallel Tempering temperature-swap probability of 0.2, and for the Hamiltonian Monte Carlo Molecular Dynamics trajectories, we used $L = 100$ steps with a step size of $\epsilon = 1.5$ femtoseconds. Empirically, we found that these parameters were sufficient to explore the potential energy landscape for a variety of molecules. Each molecule has approximately 560 conformations. There are minor differences in the total number of generated conformations for each molecule, because conformations are only proposed and then accepted with a certain probability in PT-HMC simulations.

In the fourth example of Figure 6, the left hand side of the molecule has a methylbenzene group, which is symmetric under flipping of the aromatic ring, but the ground truth distribution is an odd function. We carefully inspected our data and found that for nearly all molecules in our training data that contain relevant symmetry configurations, we properly recover even functions. We identified several with odd functions (12 in NMRShiftDB) and systematically investigated rotational barriers. We find that the energy barriers are so high (hundreds of kcal/mol) that it is possible that such 180 degree flips would never occur in nature. For this example, the energy barrier is approximately 800 kcal/mol based on the MMFF implementation in RDKit.

B. Dataset Details

For NMRShiftDB we took 32,171 molecules from NMRShiftDB (Kuhn, 2019), a popular database of nuclear magnetic resonance spectra, limiting it to molecules with up to 64 atoms of elements H, C, O, N, F, S, P, Cl and no formal charges or radicals. We ensured that every molecule had an assigned stereo chemistry.

For GDB-17 we took a random 134,228 molecule subset of the publically-available 50M "lead-like" molecules made available by the GDB-17 enumeration of chemical space (Ruddigkeit et al., 2012). As GDB provides non-isomeric SMILES strings, we leveraged RDKit's stereo enumeration code and identified at least one stereoisomer with a reasonable force-field energy (that is, no pathological steric hindrance).

In Figures 8 and 9, we examine the distribution of number of rotatable bonds and number of atoms in our NMRShiftDB and GDB-17 datasets.

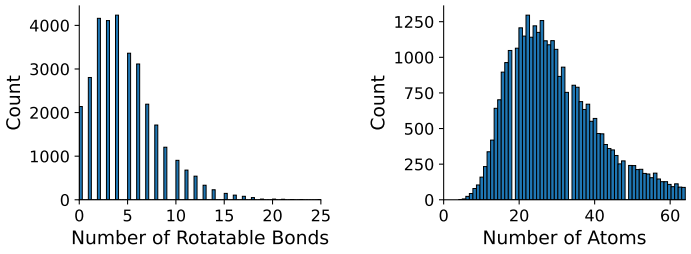
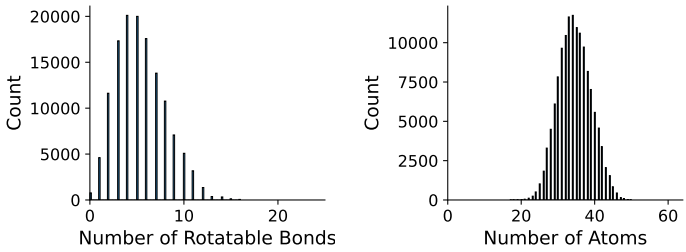


Figure 8. NMRShiftDB dataset statistics. **(a)** Distribution of number of rotatable bonds per molecule in our NMRShiftDB data. **(b)** Distribution of number of atoms per molecule in our NMRShiftDB data.



(a) (b)

Figure 9. GDB-17 dataset statistics. **(a)** Distribution of number of rotatable bonds per molecule in our GDB-17 data. **(b)** Distribution of number of atoms per molecule in our GDB-17 data.

C. Dihedral Angle Encoding for Rotatable Bonds

We use a breadth-first search approach based on the Cahn-Ingold-Prelog (CIP) (Cahn et al., 1966) rules to determine which four atoms, including the two central atoms, should be chosen to define a given rotatable bond torsion angle. For each of the two central atoms, we use the following procedure. We construct a list of neighbors excluding the other central atom. If, among this list, there is one with a unique largest atomic number, then we select that atom. If there is a tie, atoms at a distance of two bonds from the central atom are examined. We replace each atom in the original list with a list of the other atoms bonded to it that have not yet been examined. These lists are compared atom by atom, and at the earliest difference, the group with the highest atomic number is given higher priority. This process is repeated recursively until any ties are broken. In the case where there is a tie at distance n but there are no bonded atoms at distance $n + 1$, the first atom or list of atoms is given priority.

D. Approximate Independence of Rotatable Bond Torsion Angle Distributions

In this work, we make the assumption that individual rotatable bond torsion angle distributions are approximately independent. Butane is an example of a molecule that supports this assumption. Figure 10 shows an example pair of rotatable bonds in butane whose joint torsion angle distribution is approximately rank-1, which indicates independence. There are, of course, cases where pairs of rotatable bonds are not approximately rank-1, as exhibited in Figure 11. However, we show in Figure 12 that on average, across a large set of molecules, pairs of rotatable bonds have a joint distribution that is approximately rank-1, with the approximation improving when the bonds are farther apart within a molecule.

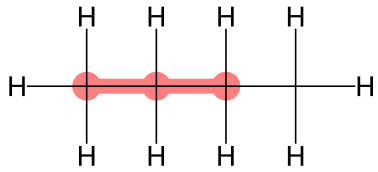
E. Mode Analysis for Rotatable Bond Torsion Angle Distributions

In this work, we make the assumption that most individual rotatable bond torsion angle distributions are multi-modal, typically with up to four distinct modes. In Figure 13, we show that for a large subset of molecules from our NMRShiftDB data, most distributions have fewer than four distinct modes.

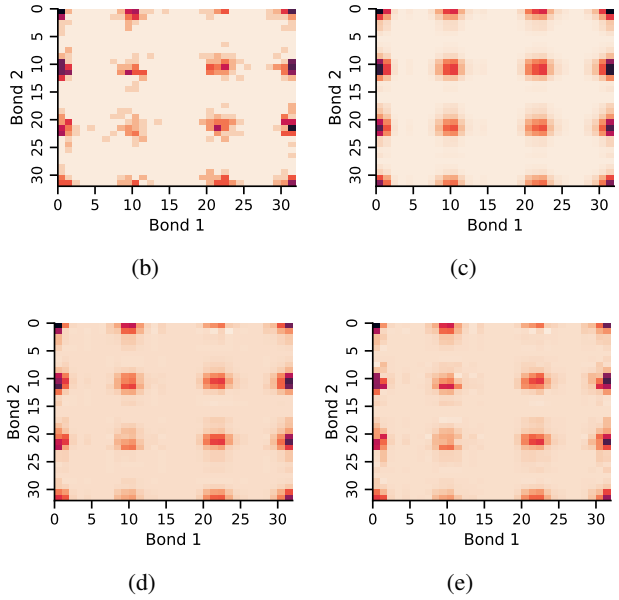
F. Molecular Graph Featurization

For each node in a molecular graph, we create a feature tensor to use in the graph neural network. These features are generated from basic molecular properties found in RDKit. They are padded with leading and/or trailing zeros so that each node has the same feature tensor shape whether it corresponds to a bond, atom or angle. In Tables 3 and 4 we describe the features used for bond and atom nodes, respectively. For angle nodes, we use a single feature. For the three atoms that correspond to an angle node, we compute the angle between them by using pairwise distances based on the average of upper and lower distance bounds provided by RDKit.

Therefore, in total, each node has a feature tensor with 15 + 113 + 1 = 129 elements (# bond + # atom + # angle). -10 is used as the sentinel value in place of features for other node types.



(a)



(b) (c)

(d) (e)

Figure 10. Rank-1 example. **(a)** Molecular graph of butane along with two adjacent rotatable bonds highlighted. **(b)** A histogram of the torsion angle joint distribution of the highlighted rotatable bonds, using 32 bins. **(c)** shows a rank-1 approximation to this histogram with reconstruction mean squared error (MSE) 0.47, **(d)** shows a rank-2 approximation with MSE 0.31, and **(e)** shows a rank-3 approximation with MSE 0.21. Although the rank-1 MSE is higher, it's clear from the visualization that the ground truth distribution is approximately rank-1, indicating independence.

G. Computing and Setting Chiralities using RDKit

To determine whether an atom has R or S chirality, we compute the oriented volume formed by the atom and its three neighbors (Ganea et al., 2021). If the oriented volume is 1, then we say the atom has R chirality, and if the oriented volume is -1, we say the atom has S chirality. The oriented volume is given by:

$$OV(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4) = sign\left(\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{vmatrix}\right) \tag{5}$$

We assign atoms to these four vectors in a consistent fashion by using the same CIP priority rules that we use for selecting

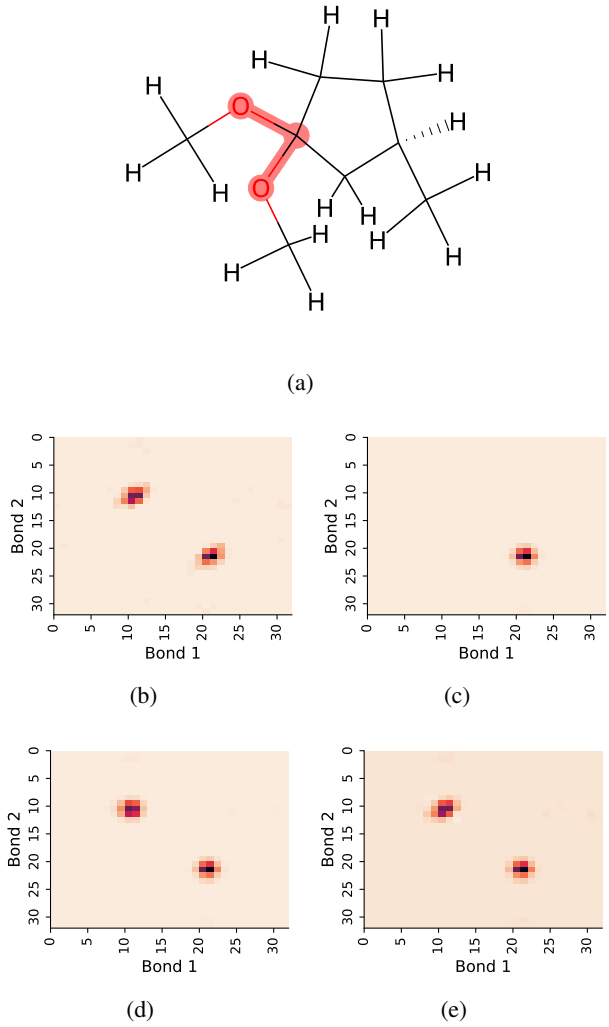


Figure 11. Non-rank-1 example. **(a)** Molecular graph of a molecule from NMRShiftDB along with two adjacent rotatable bonds highlighted. **(b)** A histogram of the torsion angle joint distribution of the highlighted rotatable bonds, using 32 bins. **(c)** shows a rank-1 approximation to this histogram with reconstruction mean squared error (MSE) 8.01, **(d)** shows a rank-2 approximation with MSE 0.53, and **(e)** shows a rank-3 approximation with MSE 0.28. This is a clear example where the joint distribution is not approximately rank-1.

the four atoms that define a rotatable bond torsion angle. The neighbors of the atom that have first, second, and third priority are assigned to $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$, respectively. The atom itself is assigned to $\mathbf{p}_4$.

If we need to flip the chirality from R to S or vice versa, we do the following. If A is the atom, we 1) compute the plane formed by A's three neighbors 2) compute the projection of A onto the plane, 3) compute the plane formed by A, one of its neighbors, and A's projection onto the neighbor plane, and 4) reflect all atoms across this plane.

H. Additional Evaluations

In this section we show additional evaluations. When evaluating without the Torsional Diffusion constraints, we use 997 out of the 1000 random molecules from NMRShiftDB-test and 997 out of the 1000 random molecules from GDB-17-test after removing molecules for which ETKDG embedding failed. In Figure 14, we evaluate rotatable bond torsion angle distributions without the Torsional Diffusion constraint. As in section 3.3, VonMisesNet and VonMisesNet-Filtered outperform the other methods. In Figures 15, 16, and 17, we evaluate 1-n pairwise distance distributions up to $n = 10$. As

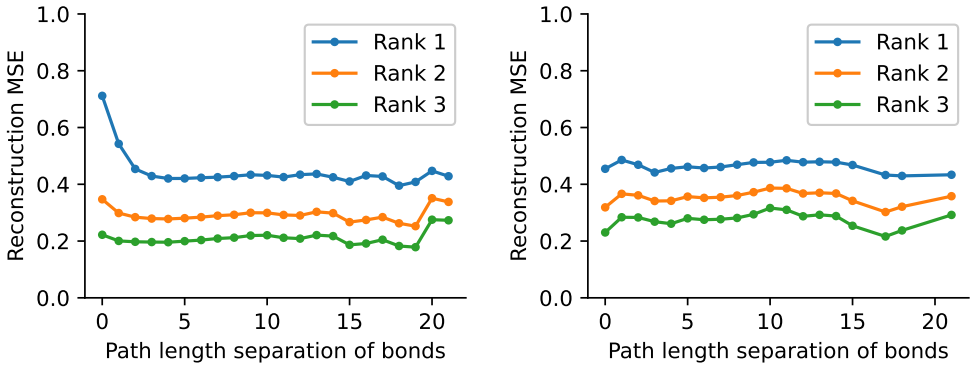


Figure 12. Large-scale rank-1 analysis. For each pair of rotatable bonds in each molecule of a random subset of 4,225 molecules from our NMRShiftDB data, we computed a 2D histogram of the torsion angle joint distribution using 32 bins. We then computed the reconstruction mean squared error (MSE) of the rank-1, rank-2, and rank-3 approximations to this histogram. We show these averages as a function of the graph distance between rotatable bond pairs. **(a)** shows that the rank-1 value rapidly converges to approximately 0.4, which is below the rank-1 value from butane in Figure 10. The average MSE value across all pairs is 0.51. This suggests that pairs of rotatable bonds have a joint distribution that is approximately rank-1, with the approximation improving when the bonds are farther apart within a molecule. **(b)** shows a similar plot that is restricted to rotatable bonds attached to methyl groups, which we would expect to be pairwise independent. The rank-1 value indicates approximate independence across all path lengths.

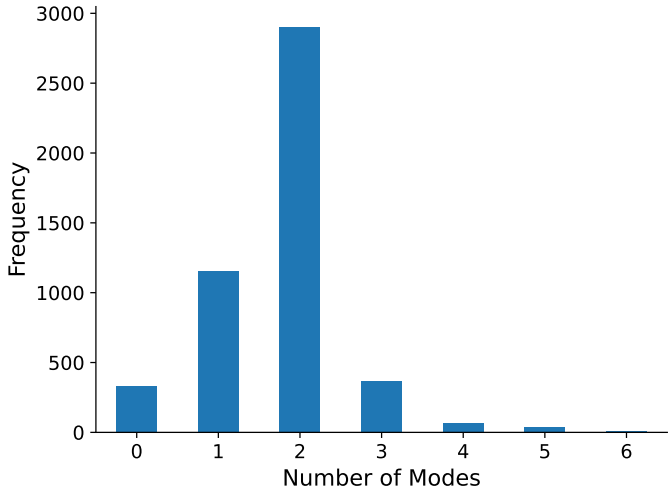


Figure 13. For each rotatable bond in a random subset of 4,870 molecules from our NMRShiftDB data, we computed the number of modes by counting maxima in a kernel density estimate of the angle distribution. The plot above shows the frequency distribution of number of modes.

in section 3.4, VonMisesNet, VonMisesNet-Filtered, and Torsional Diffusion generally outperform the other methods when restricted to rotatable bonds, shown in Figure 15. The performance of VonMisesNet-Filtered remains strong for $n \geq 6$, while the performance of VonMisesNet degrades. The performance of Torsional Diffusion tends to improve relative to VonMisesNet as a function of n. VonMisesNet and VonMisesNet-Filtered outperform the other methods on most metrics when all torsions are allowed except those belonging to non-aromatic rings, shown in Figure 16, and ETKDG-Clean performs best with no restrictions, shown in 17. In Figures 18, 19, and 20, we evaluate on the random molecules from NMRShiftDB-test and GDB-17-test without the Torsional Diffusion constraints. Similar trends hold.

Table 3. Features for Bond Nodes. Note: RDKit BondStereo types: {ANY, CIS, E, NONE, TRANS, Z}

FEATURE	DESCRIPTION	NUMBER OF ELEMENTS
BOND TYPE	ONE HOT ENCODED FROM {0, 1, 1.5, 2, 3}	5
CONJUGATED	IS BOND CONJUGATED	1
IN RING	IS BOND PART OF A RING	1
STEREO	BOND STEREO TYPE, ONE HOT ENCODED FROM RDKIT BONDSTEREO TYPES	6
SAME RING	ARE ENDPOINTS OF BOND ARE IN ANY RING TOGETHER	1
ROTATABLE BOND	IS BOND ROTATABLE	1
TOTAL		15

Table 4. Features for Atom Nodes. Note: RDKit ChiralTypes: {UNSPECIFIED, Tetrahedral CW, Tetrahedral CCW, OTHER, Tetrahedral, Allene, Square Planar, Trigonal Bipyramidal, Octahedral}. MMFF Atom Types selected: [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 37, 38, 39, 40, 42, 43, 44, 46, 48, 59, 62, 63, 64, 65, 66, 70, 71, 72, 74, 75, 78].

FEATURE	DESCRIPTION	NUMBER OF ELEMENTS
ATOMIC NUMBER	ONE HOT ENCODED FROM {H, C, O, N, F, P, S, CL}	8
VALENCE	ONE HOT ENCODED FROM 1-6	6
AROMATICITY	WHETHER ATOM IS IN AROMATIC STRUCTURE, DETERMINED BY RDKIT	1
HYBRIDIZATION	ONE HOT ENCODED FROM $\{s, sp, sp^2, sp^3, sp^3d, sp^3d^2$, UNSPECIFIED$\}$	7
PARTIAL CHARGE	GASTEIGER CHARGE FROM RDKIT (SET TO ZERO IF NOT FINITE)	1
FORMAL CHARGE	PRESENCE OF NET CHARGE, ONE HOT ENCODED FROM {-1, 0, 1}	3
COVALENT RADIUS	RDKIT COVALENT RADIUS	1
VAN DER WAALS RADIUS	RDKIT VAN DER WAALS RADIUS	1
DEFAULT VALENCE	VALENCE OF ATOM ON PERIODIC TABLE, ONE HOT ENCODED FROM 1-6	6
RINGS	WHETHER THE ATOM IS IN A RING OF SIZE N FOR N FROM 3-8	6
CHIRALITY	ONE HOT ENCODED FROM RDKIT CHIRALTYPES	9
MMFF ATOM TYPES	ATOM TYPE FROM RDKIT'S MMFFMOLPROPERTIES, ONE HOT ENCODED	51
DEGREE	ONE HOT ENCODED FROM 1-6	6
NUMBER OF HYDROGENS	TOTAL NUMBER OF HYDROGENS ON ATOM, ONE HOT ENCODED FROM 0-3	4
RADICAL ELECTRONS	NUMBER OF RADICAL ELECTRONS, ONE HOT ENCODED FROM 0-2	3
TOTAL		113

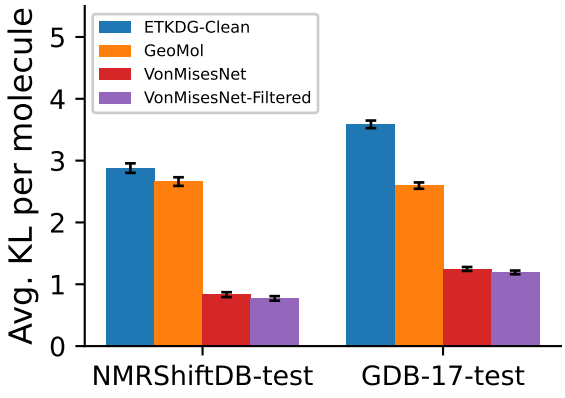
I. Enforcing Consistent Atom Indexing

When ETKDG generates multiple conformations for a single molecule, it can arbitrarily swap atom indices from one conformation to the next. This can happen specifically for atoms connected to a central atom that has four neighbors but does not have an existing chiral tag, or for neighbors of endpoint atoms of double bonds where the bond does not have a *cis*/*trans* stereochemistry tag and both endpoint atoms have more than one neighbor. Our metrics for computing pairwise distances require consistent atom indices, because we measure distances between specific indices across all conformations of a molecule. So, given an input molecule with an existing 3D geometry, we use the following procedure to prevent atom index swapping for any further conformations generated by ETKDG. For each atom that has four neighbors and no chiral tag, we compute the chirality based on the oriented volume and then assign an RDKit chirality tag accordingly. This will force future conformations generated by ETKDG to preserve the chirality and therefore preserve the current atom indexing. For each double bond where the bond does not have a *cis*/*trans* stereochemistry tag and both endpoint atoms have more than one neighbor, we set the double bond tag to *cis* in a consistent manner. This will force future conformations generated by ETKDG to preserve the stereochemistry of the double bond and therefore preserve the current atom indexing. When we generate conformations using this procedure, we refer to the method as ETKDG-Clean.

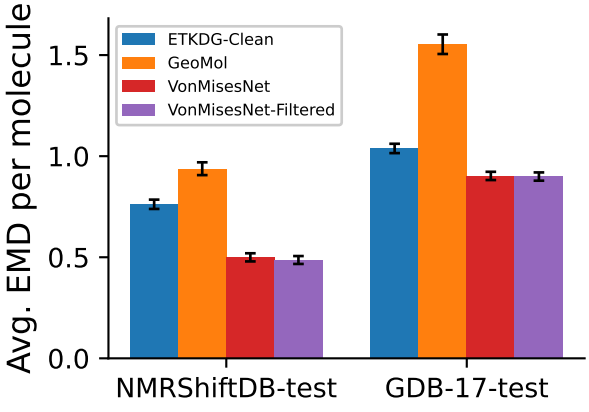
When re-training GeoMol with our datasets, we noticed that failures occurred in the form of massive spikes in the train and validation loss for some molecules. We found that this occurred when the atom indices in RDKit were ordered in such a way that did not correspond to an ordering that results from creating the molecule directly from a SMILES string representation. This suggests that GeoMol requires atom indices to have the ordering based on a SMILES parsing. Therefore, for all input molecules to GeoMol for either training or inference, we first permute the atom indices so that they correspond to a SMILES ordering. Post-inference, we undo the permutation so that the atom indexing is consistent when comparing to other conformation generators. As a precaution, we used the same reordering procedure when running inference with Torsional Diffusion.

J. VonMisesNet Architecture

VonMisesNet is a graph neural network that processes a multi-partite graph representation of molecules. The neural network architecture is as follows. We process the node features in the input graph with a linear layer followed by Layer Norm. Subsequently, a series of graph convolution layers perform message passing between the graph nodes, where each message passing step is followed by a linear layer and ReLU activation, one for each step. Message passing is performed using matrix multiplication between the graph adjacency matrix and the node feature vectors, so that the messages are the current features and the aggregation function is a simple mean over neighbors. This results in a set of hidden vectors, one for each node in the graph. The ordering of the nodes is computed in a canonical way based on the ordering of atoms and bonds given by RDKit. For each prediction task, we process a relevant subset of the nodes using a separate readout function, i.e., feedforward neural network (FFN). Each of these FFNs applies Layer Norm to the inputs, applies a linear layer with size 128 followed by the ReLU activation, and then applies a final output linear layer. There is an FFN for the bond angle predictions that processes all of the angle nodes, an FFN for the bond length predictions that processes all of the bond nodes, and an FFN for the chirality probability predictions that processes all of the chirality inversion atom nodes. The output layer for each of these FFNs has size 1, and we also take the square of the outputs in order to enforce positive values. There are three FFNs that predict the von Mises mixture mean, concentration, and weight parameters, respectively, which process all of the rotatable bond nodes that do not have a chirality inversion endpoint atom. There are also two more sets of three FFNs that predict the von Mises mixture parameters, one for conditioning on R chirality and the other for S chirality, which



(a)



(b)

Figure 14. Rotatable bond torsion angle distributions evaluation without Torsional Diffusion constraints. **(a)** Average KL divergence of rotatable bond torsion angle distributions, per molecule, relative to PT-HMC ground truth in 997 random molecules from NMRShiftDB-test and 997 random molecules from GDB-17-test. Standard error bars are shown in black. The KL divergence is measured with 32 bins. **(b)** Average EMD of rotatable bond torsion angle distributions, per molecule, relative to PT-HMC ground truth for the same sets of molecules.

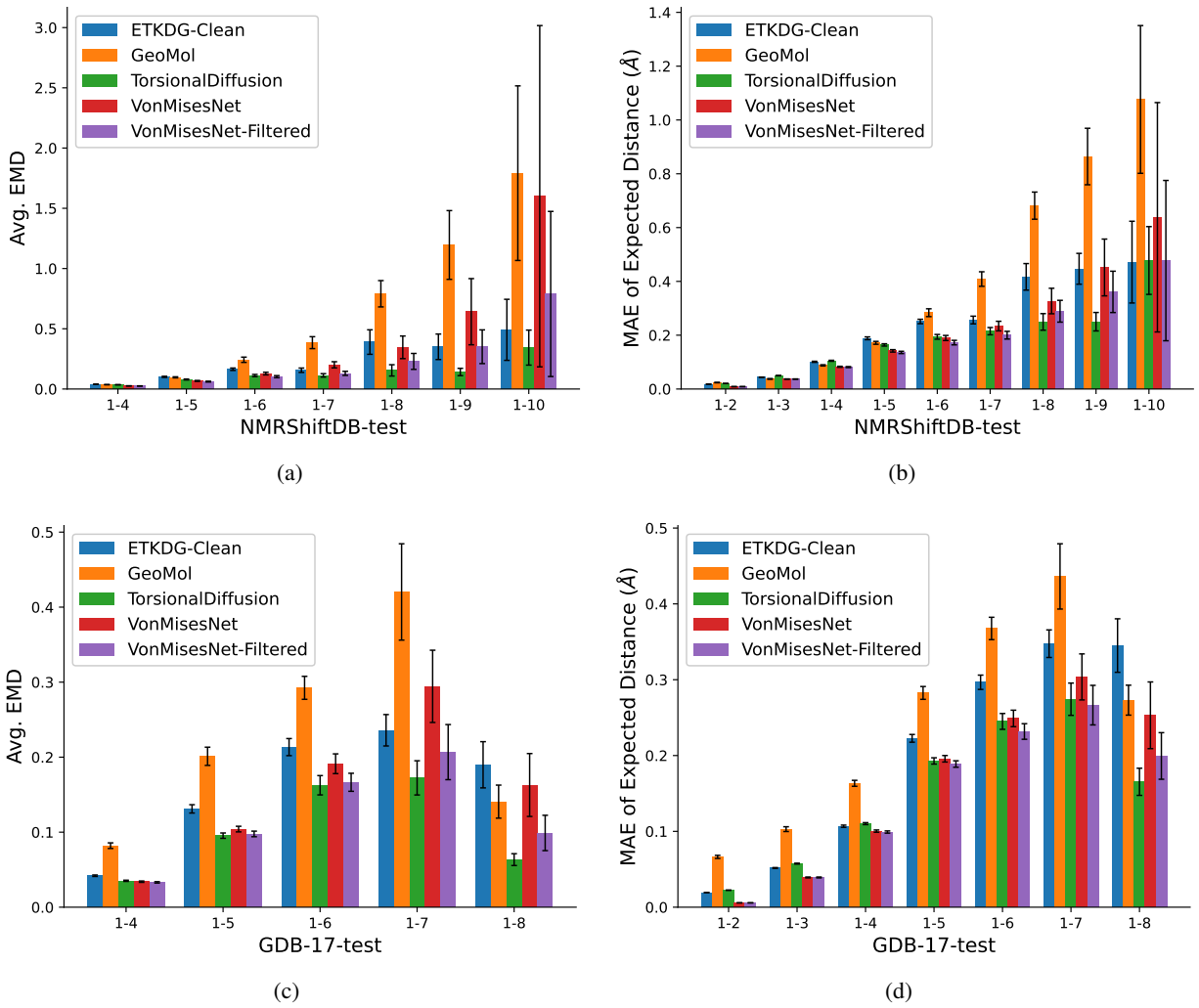


Figure 15. Pairwise distance distributions evaluation, up to 1-10, rotatable. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 538 molecules from NMRShiftDB-test and 610 molecules from GDB-17-test. We evaluate distances for which every intermediate bond along the shortest path is rotatable. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances that are not part of a ring and 1-3 distances for which at least one of the bonds is rotatable.

process all of the rotatable bond nodes that have a chirality inversion endpoint atom. The output layer for each of these FFNs has size four. We scale the concentration outputs to be between one (mininum concentration) and 20 (maximum concentration) by applying BatchNorm1d and then a sigmoid activation to the concentration outputs, and then multiplying by the maximum and adding the minimum. We apply softmax to the weights outputs so that they sum to one. We use 20 graph convolution layers, a hidden size of 256, a batch size of 32, and the Adam optimizer with a learning rate of 0.0001. We use gradient clipping for all of the model parameters with a cutoff value of 1.0. We scale the mean squared error losses for bond lengths, bond angles, and chirality probabilities by a factor of 32. Training on NMRShiftDB-train took 7.7 hours and training on GDB-17-train took 16.7 hours on a single NVIDIA GeForce RTX 2080 Ti GPU.

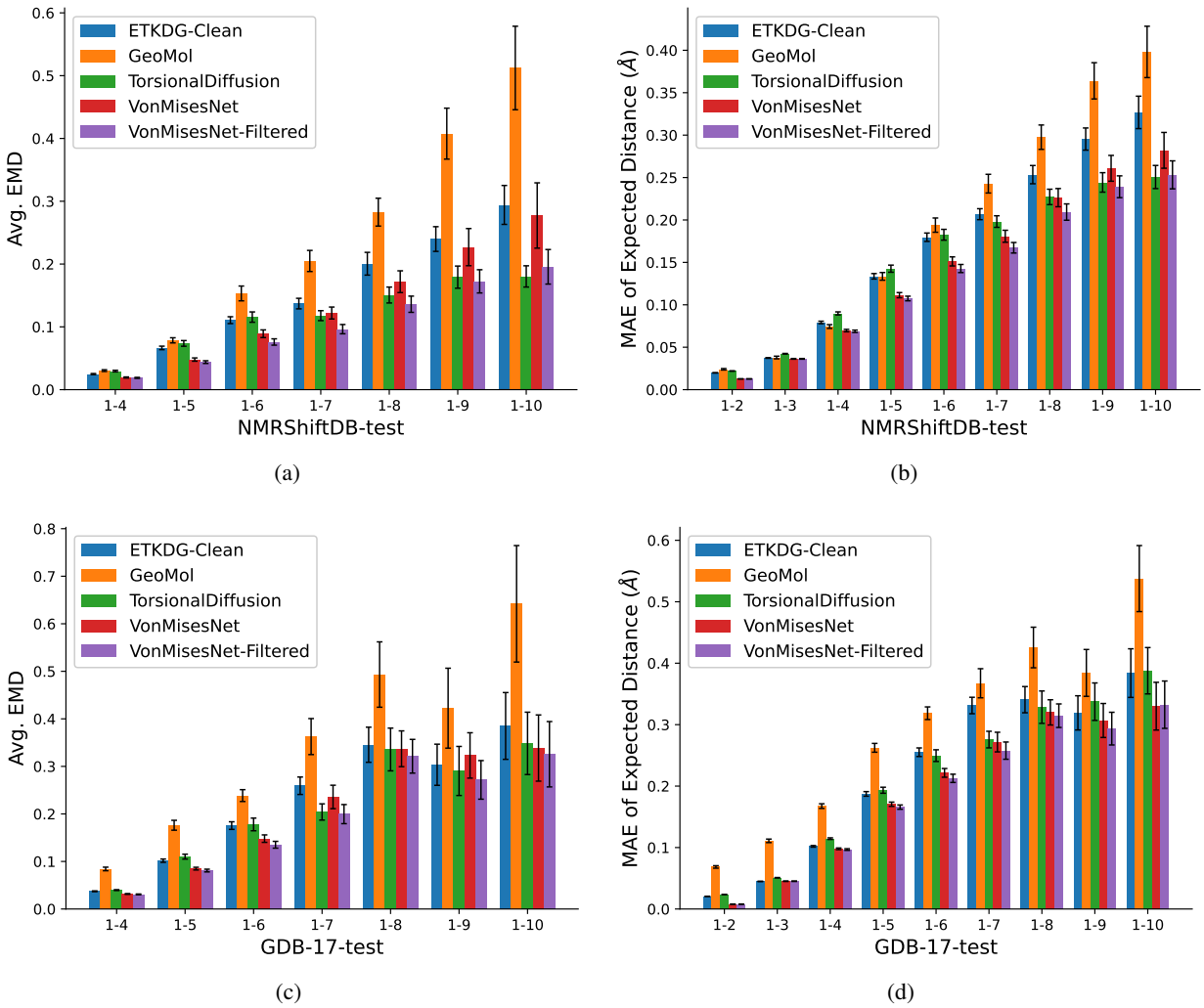


Figure 16. Pairwise distance distributions evaluation, up to 1-10, excluding non-aromatic. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 538 molecules from NMRShiftDB-test and 610 molecules from GDB-17-test. We evaluate distances for which every intermediate bond along the shortest path is not part of a non-aromatic ring. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances that are not part of a non-aromatic ring and 1-3 distances for which neither of the bonds is in a non-aromatic ring.

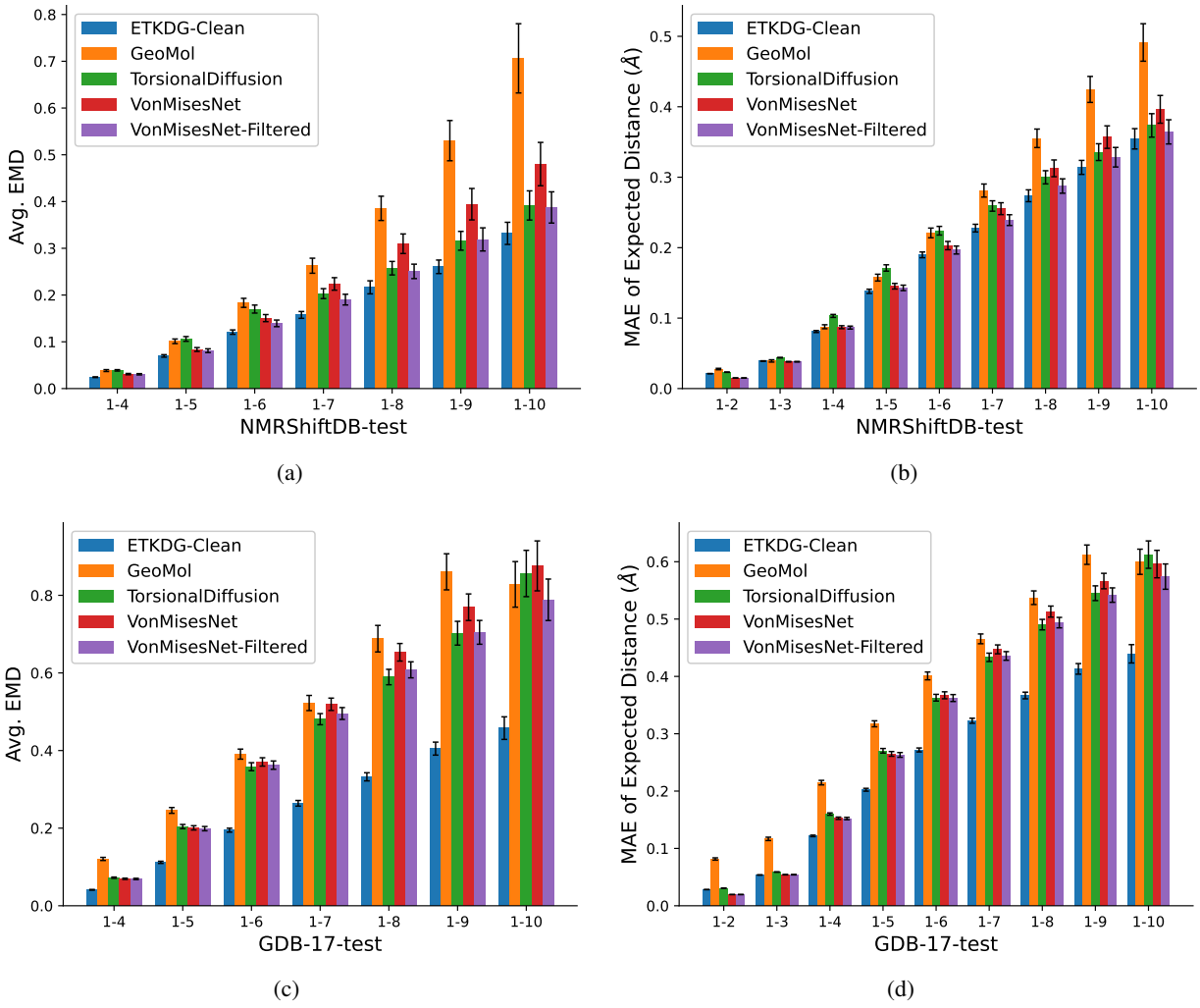


Figure 17. Pairwise distance distributions evaluation, up to 1-10, no restrictions. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 538 molecules from NMRShiftDB-test and 610 molecules from GDB-17-test. We evaluate distances without restrictions. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances and 1-3 distances without restrictions.

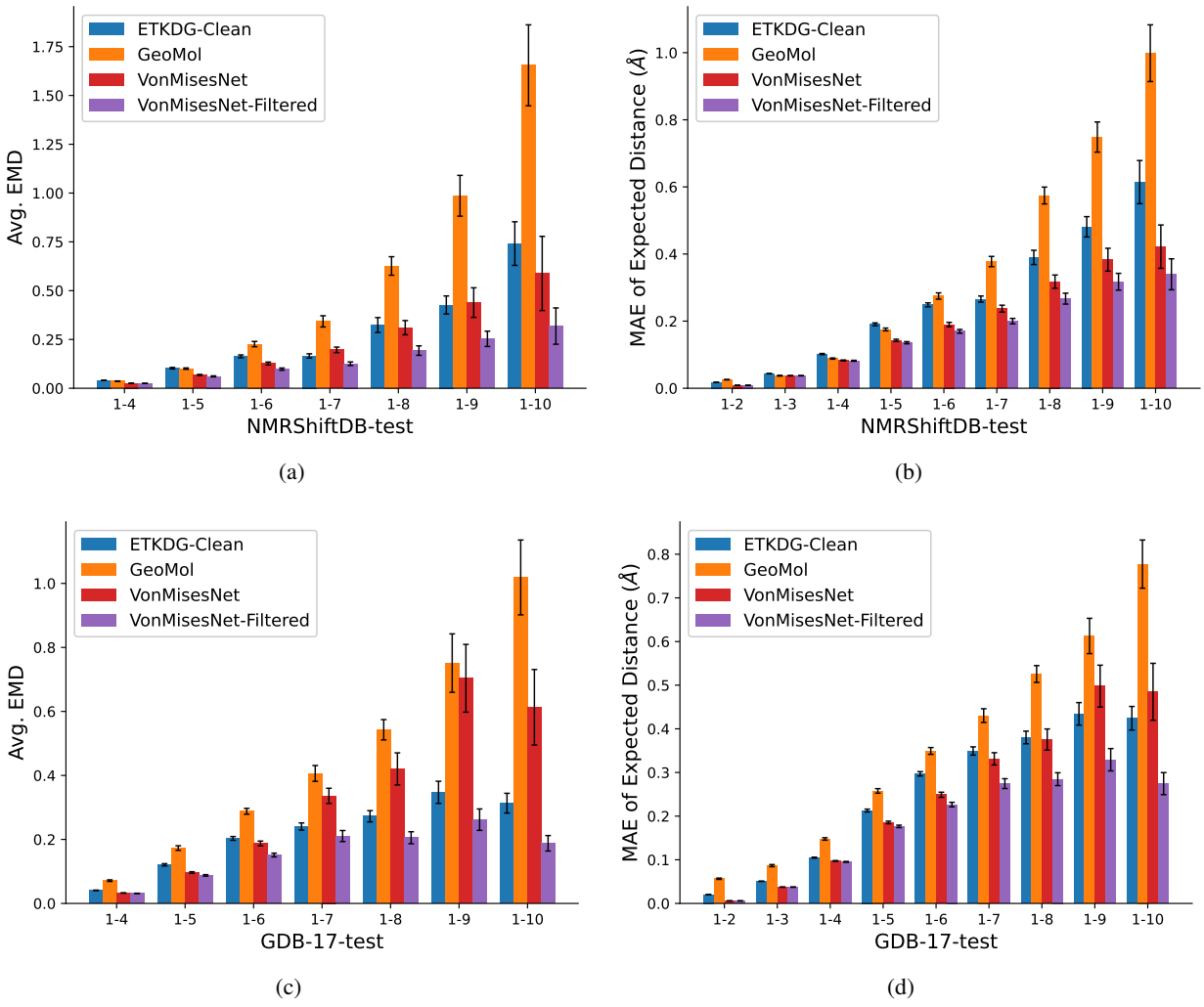


Figure 18. Pairwise distance distributions evaluation, up to 1-10, excluding Torsional Diffusion constraints, rotatable. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 997 random molecules from NMRShiftDB-test and 997 random molecules from GDB-17-test. We evaluate distances for which every intermediate bond along the shortest path is rotatable. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances that are not part of a ring and 1-3 distances for which at least one of the bonds is rotatable.

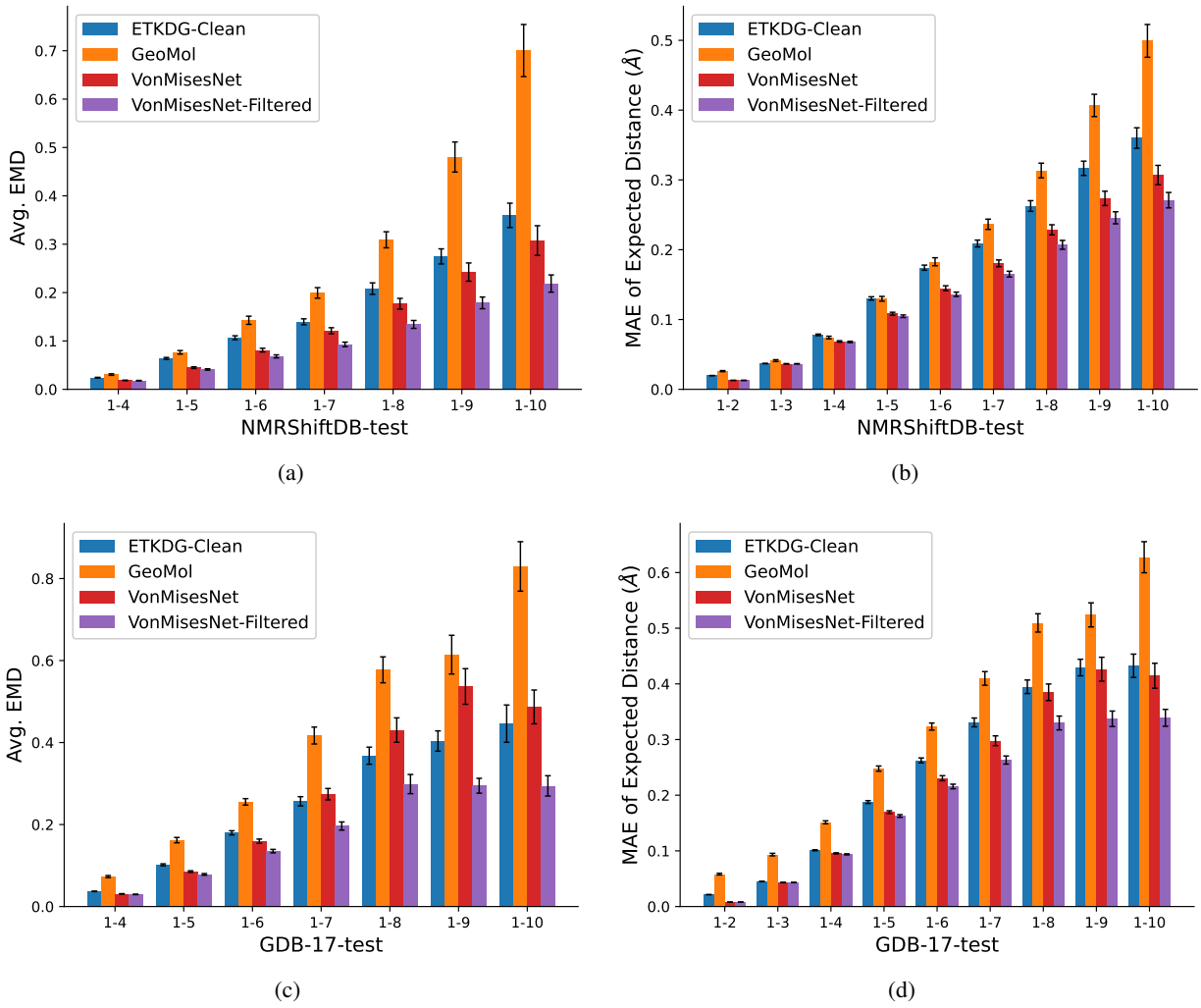


Figure 19. Pairwise distance distributions evaluation, up to 1-10, excluding Torsional Diffusion constraints, excluding non-aromatic. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 997 random molecules from NMRShiftDB-test and 997 random molecules from GDB-17-test. We evaluate distances for which every intermediate bond along the shortest path is not part of a non-aromatic ring. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances that are not part of a non-aromatic ring and 1-3 distances for which neither of the bonds is in a non-aromatic ring.

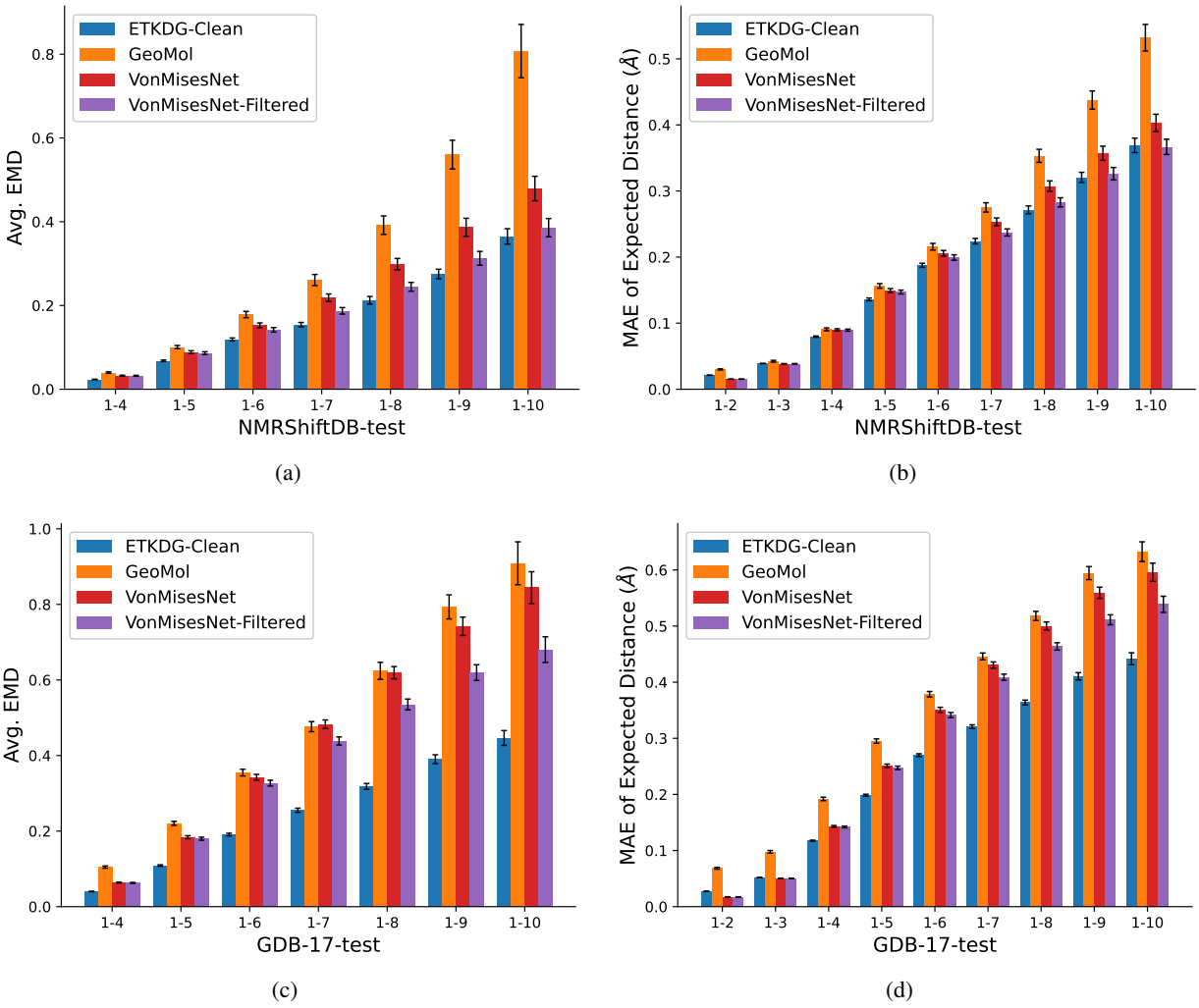
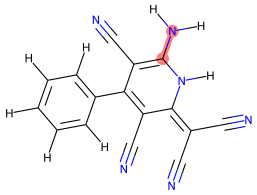
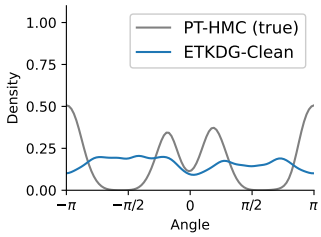


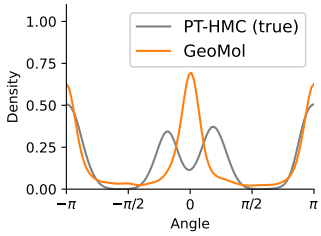
Figure 20. Pairwise distance distributions evaluation, up to 1-10, excluding Torsional Diffusion constraints, no restrictions. We evaluate pairwise distance distributions relative to PT-HMC ground truth for 997 random molecules from NMRShiftDB-test and 997 random molecules from GDB-17-test. We evaluate distances without restrictions. In **(a)** and **(c)** we compare the average EMD, per molecule, and in **(b)** and **(d)** we compare the MAE of the expected distance, per molecule. For the expected distance evaluations, we additionally include 1-2 distances and 1-3 distances without restrictions.



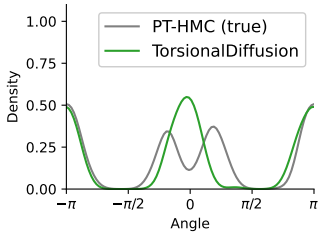
(a)



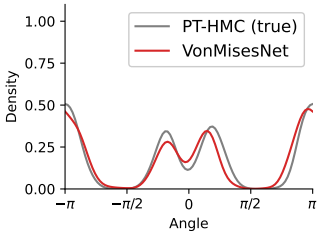
(b)



(c)



(d)



(e)

Figure 21. Predicting chirality inversion example, extended. **(a)** A molecule from NMRShiftDB-test, where the highlighted rotatable bond is between a carbon atom and a chirality inversion nitrogen atom. Kernel density estimates of the rotatable bond torsion angle from PT-HMC and from ETKDG-Clean, GeoMol, Torsional Diffusion, and VonMisesNet are shown in **(b)**, **(c)**, **(d)**, and **(e)**, respectively.