

# Learning Gradient Fields for Molecular Conformation Generation

Chence Shi<sup>\*12</sup> Shitong Luo<sup>\*3</sup> Minkai Xu<sup>12</sup> Jian Tang<sup>145</sup>

## Abstract

We study a fundamental problem in computational chemistry known as molecular conformation generation, trying to predict stable 3D structures from 2D molecular graphs. Existing machine learning approaches usually first predict distances between atoms and then generate a 3D structure satisfying the distances, where noise in predicted distances may induce extra errors during 3D coordinate generation. **Inspired by the traditional force field methods for molecular dynamics simulation, in this paper, we propose a novel approach called ConfGF by directly estimating the gradient fields of the log density of atomic coordinates.** The estimated gradient fields allow directly generating stable conformations via Langevin dynamics. However, the problem is very challenging as the gradient fields are roto-translation equivariant. We notice that estimating the gradient fields of atomic coordinates can be translated to estimating the gradient fields of interatomic distances, and hence develop a novel algorithm based on recent score-based generative models to effectively estimate these gradients. Experimental results across multiple tasks show that ConfGF outperforms previous state-of-the-art baselines by a significant margin. The code is available at <https://github.com/DeepGraphLearning/ConfGF>.

## 1. Introduction

Graph-based representations of molecules, where nodes represent atoms and edges represent bonds, have been widely applied to a variety of molecular modeling tasks, ranging from property prediction (Gilmer et al., 2017; Hu et al., 2020) to molecule generation (Jin et al., 2018; You et al., 2018; Shi et al., 2020; Xie et al., 2021). However, in reality

<sup>\*</sup>Equal contribution <sup>1</sup>Mila - Quebec AI Institute, Montréal, Canada <sup>2</sup>University of Montréal, Montréal, Canada <sup>3</sup>Peking University <sup>4</sup>CIFAR AI Research Chair <sup>5</sup>HEC Montréal, Montréal, Canada. Correspondence to: Jian Tang <jian.tang@hec.ca>.

a more natural and intrinsic representation for molecules is their three-dimensional structures, where atoms are represented by 3D coordinates. These structures are also widely known as **molecular geometry or conformation**.

Nevertheless, obtaining valid and stable conformations for a molecule remains a long-standing challenge. Existing approaches to molecular conformation generation mainly rely on molecular dynamics (MD) (De Vivo et al., 2016), where coordinates of atoms are sequentially updated based on the forces acting on each atom. The per-atom forces come either from computationally intensive Density Functional Theory (DFT) (Parr, 1989) or from hand-designed force fields (Rappé et al., 1992; Halgren, 1996) which are often crude approximations of the actual molecular energies (Kanal et al., 2017). Recently, there are growing interests in developing machine learning methods for conformation generation (Mansimov et al., 2019; Simm & Hernandez-Lobato, 2020; Xu et al., 2021). In general, these approaches (Simm & Hernandez-Lobato, 2020; Xu et al., 2021) are divided into two stages, which first predict the molecular distance geometry, i.e., interatomic distances, and then convert the distances to 3D coordinates via post-processing algorithms (Liberti et al., 2014). Such methods effectively model the roto-translation equivariance of molecular conformations and have achieved impressive performance. Nevertheless, such approaches generate conformations in a two-stage fashion, where noise in generated distances may affect 3D coordinate reconstruction, often leading to less accurate or even erroneous structures. Therefore, we are seeking for an approach that is able to generate molecular conformations within a single stage.

Inspired by the traditional force field methods for molecular dynamics simulation (Frenkel & Smit, 1996), in this paper, we propose a novel and principled approach called ConfGF for molecular conformation generation. The basic idea is to directly learn the gradient fields of the **log density w.r.t. the atomic coordinates**, which can be viewed as pseudo-forces acting on each atom. With the **estimated gradient fields**, **conformations can be generated directly using Langevin dynamics** (Welling & Teh, 2011), which amounts to moving atoms toward the high-likelihood regions guided by pseudo-forces. The key challenge with this approach is that such gradient fields of atomic coordinates are roto-translation equivariant, i.e., the **gradients rotate together with the molecular system**

and are invariant under translation.

To tackle this challenge, we observe that the interatomic distances are continuously differentiable w.r.t. atomic coordinates. Therefore, we propose to first estimate the gradient fields of the log density w.r.t. interatomic distances, and show that under a minor assumption, the gradient fields of the log density of atomic coordinates can be calculated from these w.r.t. distances in closed form using chain rule. We also demonstrate that such designed gradient fields satisfy roto-translation equivariance. In contrast to distance-based methods (Simm & Hernandez-Lobato, 2020; Xu et al., 2021), our approach generates conformations in an one-stage fashion, i.e., the sampling process is done directly in the coordinate space, which avoids the errors induced by post-processing and greatly enhances generation quality. In addition, our approach requires no surrogate losses and allows flexible network architectures, making it more capable to generate diverse and realistic conformations.

We conduct extensive experiments on GEOM (Axelrod & Gomez-Bombarelli, 2020) and ISO17 (Simm & Hernandez-Lobato, 2020) benchmarks, and compare ConfGF against previous state-of-the-art neural-methods as well as the empirical method on multiple tasks ranging from conformation generation and distance modeling to property prediction. Numerical results show that ConfGF outperforms previous state-of-the-art baselines by a clear margin.

## 2. Related Work

**Molecular Conformation Generation** Existing works on molecular conformation generation mainly rely on molecular dynamics (MD) (De Vivo et al., 2016). Starting from an initial conformation and a physical model for interatomic potentials (Parr, 1989; Rappé et al., 1992), the conformation is sequentially updated based on the forces acting on each atom. However, these methods are inefficient, especially when molecules are large as they involve computationally expensive quantum mechanical calculations (Parr, 1989; Shim & MacKerell Jr, 2011; Ballard et al., 2015). Another much more efficient but less accurate sort of approach is the empirical method, which fixes distances and angles between atoms in a molecule to idealized values according to a set of rules (Blaney & Dixon, 2007).

Recent advances in deep generative models open the door to data-driven approaches, which have shown great potential to strike a good balance between computational efficiency and accuracy. For example, Mansimov et al. (2019) introduce the Conditional Variational Graph Auto-Encoder (CVGAE) which takes molecular graphs as input and directly generates 3D coordinates for each atom. This method does not manage to model the roto-translation equivariance of molecular conformations, resulting in great differences between

generated structures and ground-truth structures. To address the issue, Simm & Hernandez-Lobato (2020) and Xu et al. (2021) propose to model interatomic distances of molecular conformations using VAE and Continuous Flow respectively. These approaches preserve roto-translation equivariance of molecular conformations and consist of two stages — first generating distances between atoms and then feeding the distances to the distance geometry algorithms (Liberti et al., 2014) to reconstruct 3D coordinates. However, the distance geometry algorithms are vulnerable to noise in generated distances, which may lead to inaccurate or even erroneous structures. Another line of research (Gogineni et al., 2020) leverages reinforcement learning for conformation generation by sequentially determining torsional angles, which relies on an additional classical force field for state transition and reward evaluation. Nevertheless, it is incapable of modeling other geometric properties such as bond angles and bond lengths, which distinguishes it from all other works.

**Neural Force Fields** Neural networks have also been employed to estimate molecular potential energies and force fields (Schütt et al., 2017; Zhang et al., 2018; Hu et al., 2021). The goal of these models is to predict energies and forces as accurate as possible, serving as an alternative to quantum chemical methods such as Density Function Theory (DFT) method. Training these models requires access to molecular dynamics trajectories along with ground-truth energies and forces from expensive quantum mechanical calculations. Different from these methods, our goal is to *generate equilibrium conformations within a single stage*. To this end, we define gradient fields<sup>1</sup> analogous to force fields, which serve as pseudo-forces acting on each atom that gradually move atoms toward the high-density areas until they converges to an equilibrium. The only data we need to train the model is equilibrium conformations.

## 3. Preliminaries

### 3.1. Problem Definition

**Notations** In this paper, a molecular graph is represented as an undirected graph  $G = (V, E)$ , where  $V = \{v_1, v_2, \dots, v_{|V|}\}$  is the set of nodes representing atoms, and  $E = \{e_{ij} \mid (i, j) \subseteq V \times V\}$  is the set of edges between atoms in the molecule. Each node  $v_i \in V$  is associated with a nuclear charge  $Z_i$  and a 3D vector  $\mathbf{r}_i \in \mathbb{R}^3$ , indicating its atomic type and atomic coordinate respectively. Each edge  $e_{ij} \in E$  is associated with a bond type and a scalar  $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|_2$  denoting the Euclidean distance between the positions of  $v_i$  and  $v_j$ . All distances between connected nodes can be represented as a vector  $\mathbf{d} = \{d_{ij}\} \in \mathbb{R}^{|E|}$ .

As the bonded edges in a molecule are not sufficient to char-

<sup>1</sup>Such gradient fields, however, are not force fields as they are not energy conserving.

acterize a conformation (**bonds are rotatable**), we extend the original molecular graph by adding auxiliary edges, i.e., virtual bonds, between atoms that are 2 or 3 hops away from each other, which is a widely-used technique in previous work (Simm & Hernandez-Lobato, 2020; Xu et al., 2021) for reducing the degrees of freedom in 3D coordinates. The extra edges between second neighbors help fix angles between atoms, while those between third neighbors fix dihedral angles. Hereafter, we assume all molecular graphs are extended unless stated.

**Problem Definition** Given an extended molecular graph  $G = (V, E)$ , our goal is to learn a generative model that generates molecular conformations  $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{|V|}) \in \mathbb{R}^{|V| \times 3}$  based on the molecular graph.

### 3.2. Score-Based Generative Modeling

The *score function*  $s(\mathbf{x})$  of a continuously differentiable probability density  $p(\mathbf{x})$  is defined as  $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ . Score-based generative modeling (Song & Ermon, 2019; Song et al., 2020; Meng et al., 2020; Song & Ermon, 2020) is a class of generative models that approximate the score function of  $p(\mathbf{x})$  using neural networks and generate new samples with Langevin dynamics (Welling & Teh, 2011). Modeling the score function instead of the density function allows getting rid of the intractable partition function of  $p(\mathbf{x})$  and avoiding extra computation on getting the higher-order gradients of an energy-based model (Song & Ermon, 2019).

To address the difficulty of estimating score functions in the regions without training data, Song & Ermon (2019) propose to perturb the data using Gaussian noise of various intensities and jointly estimate the score functions, i.e.,  $s_{\theta}(\mathbf{x}; \sigma_i)$ , for all noise levels. In specific, given a data distribution  $p(\mathbf{x})$  and a sequence of noise levels  $\{\sigma_i\}_{i=1}^L$  with a noise distribution  $q_{\sigma_i}(\tilde{\mathbf{x}} | \mathbf{x})$ , e.g.,  $\mathcal{N}(\tilde{\mathbf{x}} | \mathbf{x}, \sigma_i^2 \mathbf{I})$ , the training objective  $\ell(\theta; \sigma_i)$  for each noise level  $\sigma_i$  is as follows:

$$\frac{1}{2} \mathbb{E}_{p(\mathbf{x})} \mathbb{E}_{q_{\sigma_i}(\tilde{\mathbf{x}} | \mathbf{x})} [\|s_{\theta}(\tilde{\mathbf{x}}, \sigma_i) - \nabla_{\tilde{\mathbf{x}}} \log q_{\sigma_i}(\tilde{\mathbf{x}} | \mathbf{x})\|_2^2]. \quad (1)$$

After the noise conditional score networks  $s_{\theta}(\mathbf{x}, \sigma_i)$  are trained, samples are generated using annealed Langevin dynamics (Song & Ermon, 2019), where samples from each noise level serve as initializations for Langevin dynamics of the next noise level. For detailed design choices of noise levels and sampling hyperparameters, we refer readers to Song & Ermon (2020).

### 3.3. Equivariance in Molecular Geometry

*Equivariance* is ubiquitous in physical systems, e.g., 3D roto-translation equivariance of molecular conformations or point clouds (Thomas et al., 2018; Weiler et al., 2018; Chmiela et al., 2019; Fuchs et al., 2020; Miller et al., 2020;

Simm et al., 2021). Endowing model with such inductive biases is critical for better generalization and successful learning (Köhler et al., 2020). For molecular geometry modeling, such equivariance is guaranteed by either taking gradients of a predicted invariant scalar energy (Klicpera et al., 2020; Schütt et al., 2017) or using equivariant networks (Fuchs et al., 2020; Satorras et al., 2021). Formally, a function  $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$  being equivariant can be represented as follows:<sup>2</sup>

$$\mathcal{F} \circ \rho(x) = \rho \circ \mathcal{F}(x), \quad (2)$$

where  $\rho$  is a transformation function, e.g., rotation. Intuitively, Eq. 2 says that applying the  $\rho$  on the input has the same effect as applying it to the output.

In this paper, we aim to model the score function of  $p(\mathbf{R} | G)$ , i.e.,  $s(\mathbf{R}) = \nabla_{\mathbf{R}} \log p(\mathbf{R} | G)$ . As  $\log p(\mathbf{R} | G)$  is roto-translation invariant with respect to conformations, the score function  $s(\mathbf{R})$  is therefore roto-translation equivariant with respect to conformations. To pursue a generalizable and elegant approach, we explicitly build such equivariance into the model design.

## 4. Proposed Method

### 4.1. Overview

Our approach extends the denoising score matching (Vincent, 2011; Song & Ermon, 2019) to molecular conformation generation by estimating the gradient fields of the log density of atomic coordinates, i.e.,  $\nabla_{\mathbf{R}} \log p(\mathbf{R} | G)$ . Directly parameterizing such gradient fields using vanilla Graph Neural Networks (GNNs) (Duvenaud et al., 2015; Kearnes et al., 2016; Kipf & Welling, 2016; Gilmer et al., 2017; Schütt et al., 2017; Schütt et al., 2017) is problematic, as these GNNs only leverage the node, edge or distance information, resulting in node representations being roto-translation invariant. However, the gradient fields of atomic coordinates should be roto-translation equivariant (see Section 3.3), i.e., the vector fields should rotate together with the molecular system and be invariant under translation.

To tackle this issue, we assume that the log density of atomic coordinates  $\mathbf{R}$  given the molecular graph  $G$  can be parameterized as  $\log p_{\theta}(\mathbf{R} | G) \triangleq f_G \circ g_G(\mathbf{R}) = f_G(\mathbf{d})$  up to a constant, where  $g_G : \mathbb{R}^{|V| \times 3} \rightarrow \mathbb{R}^{|E|}$  denotes a function that maps a set of atomic coordinates to a set of interatomic distances, and  $f_G : \mathbb{R}^{|E|} \rightarrow \mathbb{R}$  is a graph neural network that estimates the negative energy of a molecule based on the interatomic distances  $\mathbf{d}$  and the graph representation  $G$ . Such design is a common practice in existing literature (Schütt et al., 2017; Hu et al., 2020; Klicpera et al., 2020), which is favored as it ensures several physical in-

<sup>2</sup>Strictly speaking,  $\rho$  does not have to be the same on both sides of the equation, as long as it is a representation of the same group. Note that *invariance* is a special case of equivariance.

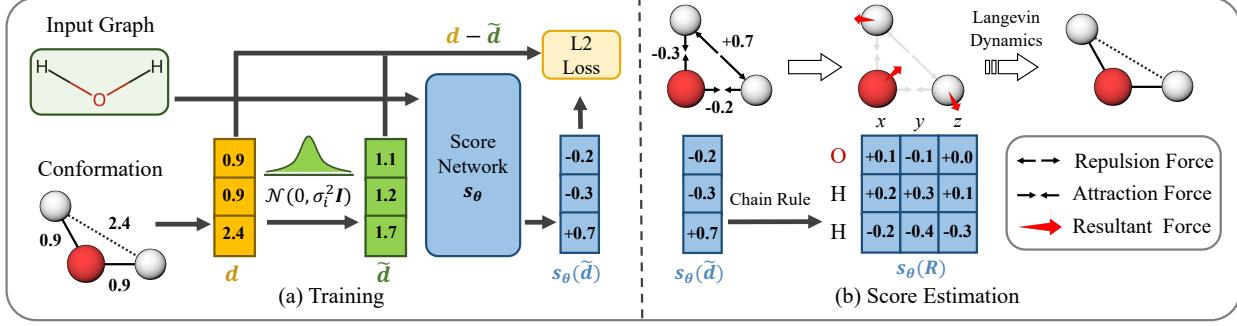


Figure 1. Overview of the proposed ConfGF approach. (a) Illustration of the training procedure. We perturb the interatomic distances with random Gaussian noise of various magnitudes, and train the noise conditional score network with denoising score matching. (b) Score estimation of atomic coordinates via chain rule. The procedure amounts to calculating the resultant force from a set of interatomic forces.

variances, e.g., rotation and translation invariance of energy prediction. We observe that the interatomic distances  $\mathbf{d}$  are continuously differentiable w.r.t. atomic coordinates  $\mathbf{R}$ . Therefore, the two gradient fields,  $\nabla_{\mathbf{R}} \log p_{\theta}(\mathbf{R} | G)$  and  $\nabla_{\mathbf{d}} \log p_{\theta}(\mathbf{d} | G)$ , are connected by chain rule, and the gradients can be backpropagated from  $\mathbf{d}$  to  $\mathbf{R}$  as follows:

$$\begin{aligned} \forall i, \mathbf{s}_{\theta}(\mathbf{R})_i &= \frac{\partial f_G(\mathbf{d})}{\partial \mathbf{r}_i} = \sum_{(i,j), e_{ij} \in E} \frac{\partial f_G(\mathbf{d})}{\partial d_{ij}} \cdot \frac{\partial d_{ij}}{\partial \mathbf{r}_i} \\ &= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \frac{\partial f_G(\mathbf{d})}{\partial d_{ij}} \cdot (\mathbf{r}_i - \mathbf{r}_j) \\ &= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \mathbf{s}_{\theta}(\mathbf{d})_{ij} \cdot (\mathbf{r}_i - \mathbf{r}_j), \end{aligned} \quad (3)$$

where  $\mathbf{s}_{\theta}(\mathbf{R})_i = \nabla_{\mathbf{r}_i} \log p_{\theta}(\mathbf{R} | G)$  and  $\mathbf{s}_{\theta}(\mathbf{d})_{ij} = \nabla_{d_{ij}} \log p_{\theta}(\mathbf{d} | G)$ . Here  $N(i)$  denotes  $i^{th}$  atom's neighbors. The second equation of Eq. 3 holds because of the chain rule over the partial derivative of the composite function  $f_G(\mathbf{d}) = f_G \circ g_G(\mathbf{R})$ .

Inspired by the above observations, we take the interatomic pairwise distances as intermediate variables and propose to first estimate the gradient fields of interatomic distances corresponding to different noise levels, i.e., training a noise conditional score network  $\mathbf{s}_{\theta}(\mathbf{d}, \sigma)$  to approximate  $\nabla_{\tilde{\mathbf{d}}} \log q_{\sigma}(\tilde{\mathbf{d}} | G)$ . We then calculate the gradient fields of atomic coordinates, i.e.,  $\mathbf{s}_{\theta}(\mathbf{R}, \sigma)$ , from  $\mathbf{s}_{\theta}(\mathbf{d}, \sigma)$  via Eq. 3, without relying on any extra parameters. Such designed score network  $\mathbf{s}_{\theta}(\mathbf{R}, \sigma)$  enjoys the property of roto-translation equivariance. Formally, we have:

**Proposition 1.** Under the assumption that we parameterize  $\log p_{\theta}(\mathbf{R} | G)$  as a function of interatomic distances  $\mathbf{d}$  and molecular graph representation  $G$ , the score network defined in Eq. 3 is roto-translation equivariant, i.e., the gradients rotate together with the molecule system and are invariant under translation.

*Proof Sketch.* The score network of distances  $\mathbf{s}_{\theta}(\mathbf{d})$  is roto-translation invariant as it only depends on interatomic distances, and the vector  $(\mathbf{r}_i - \mathbf{r}_j)/d_{ij}$  rotates together with the molecule system, which endows the score network of atomic coordinates  $\mathbf{s}_{\theta}(\mathbf{R})$  with roto-translation equivariance. See Supplementary material for a formal proof.  $\square$

After training the score network, based on the estimated gradient fields of atomic coordinates corresponding to different noise levels, conformations are generated by annealed Langevin dynamics (Song & Ermon, 2019), combining information from all noise levels. The overview of ConfGF is illustrated in Figure 1 and Figure 2. Below we describe the design of the noise conditional score network in Section 4.2, introduce the generation procedure in Section 4.3, and show how the designed gradient fields are connected with the classical laws of mechanics in Section 4.4.

## 4.2. Noise Conditional Score Networks for Distances

Let  $\{\sigma_i\}_{i=1}^L$  be a positive geometric progression with common ratio  $\gamma$ , i.e.,  $\sigma_i/\sigma_{i-1} = \gamma$ . We define the perturbed conditional distance distribution  $q_{\sigma}(\tilde{\mathbf{d}} | G)$  to be  $\int p(\mathbf{d} | G) \mathcal{N}(\tilde{\mathbf{d}} | \mathbf{d}, \sigma_i^2 \mathbf{I}) d\mathbf{d}$ . In this setting, we aim to learn a conditional score network to jointly estimate the scores of all perturbed distance distributions, i.e.,  $\forall \sigma \in \{\sigma_i\}_{i=1}^L : \mathbf{s}_{\theta}(\tilde{\mathbf{d}}, \sigma) \approx \nabla_{\tilde{\mathbf{d}}} \log q_{\sigma}(\tilde{\mathbf{d}} | G)$ . Note that  $\mathbf{s}_{\theta}(\tilde{\mathbf{d}}, \sigma) \in \mathbb{R}^{|E|}$ , we therefore formulate it as an edge regression problem.

Given a molecular graph  $G = (V, E)$  with a set of interatomic distances  $\mathbf{d} \in \mathbb{R}^{|E|}$  computed from its conformation  $\mathbf{R} \in \mathbb{R}^{|V| \times 3}$ , we first embed the node attributes and the edge attributes into low dimensional feature spaces using feedforward networks:

$$\begin{aligned} \mathbf{h}_i^0 &= \text{MLP}(Z_i), \quad \forall v_i \in V, \\ \mathbf{h}_{e_{ij}} &= \text{MLP}(e_{ij}, d_{ij}), \quad \forall e_{ij} \in E. \end{aligned} \quad (4)$$

We then adopt a variant of Graph Isomorphism Network (GINs) (Xu et al., 2019; Hu et al., 2020) to compute node

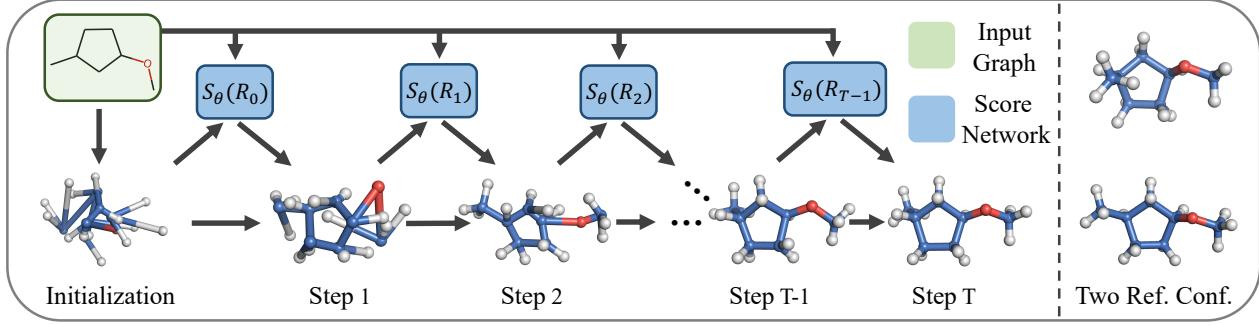


Figure 2. Generation procedure of the proposed ConfGF via Langevin dynamics. Starting from a random initialization, the conformation is sequentially updated with the gradient information of atomic coordinates calculated from the score network.

embeddings based on graph structures. At each layer, node embeddings are updated by aggregating messages from neighboring nodes and edges:

$$\mathbf{h}_i^l = \text{MLP}(\mathbf{h}_i^{l-1} + \sum_{j \in N(i)} \text{ReLU}(\mathbf{h}_j^{l-1} + \mathbf{h}_{e_{ij}})), \quad (5)$$

where  $N(i)$  denotes  $i^{\text{th}}$  atom's neighbors. After  $N$  rounds of message passing, we compute the final edge embeddings by concatenating the corresponding node embeddings for each edge as follows:

$$\mathbf{h}_{e_{ij}}^o = \mathbf{h}_i^N \parallel \mathbf{h}_j^N \parallel \mathbf{h}_{e_{ij}}, \quad (6)$$

where  $\parallel$  denotes the vector concatenation operation, and  $\mathbf{h}_{e_{ij}}^o$  denotes the final embeddings of edge  $e_{ij} \in E$ . Following the suggestions of Song & Ermon (2020), we parameterize the noise conditional network with  $\mathbf{s}_\theta(\tilde{\mathbf{d}}, \sigma) = \mathbf{s}_\theta(\tilde{\mathbf{d}})/\sigma$  as follows:

$$\mathbf{s}_\theta(\tilde{\mathbf{d}})_{ij} = \text{MLP}(\mathbf{h}_{e_{ij}}^o), \quad \forall e_{ij} \in E, \quad (7)$$

where  $\mathbf{s}_\theta(\tilde{\mathbf{d}})$  is an unconditional score network, and the MLP maps edge embeddings to a scalar for each  $e_{ij} \in E$ . Note that  $\mathbf{s}_\theta(\tilde{\mathbf{d}}, \sigma)$  is roto-translation invariant as we only use interatomic distances. Since  $\nabla_{\tilde{\mathbf{d}}} \log q_\sigma(\tilde{\mathbf{d}} \mid \mathbf{d}, G) = -(\tilde{\mathbf{d}} - \mathbf{d})/\sigma^2$ , the training loss of  $\mathbf{s}_\theta(\tilde{\mathbf{d}}, \sigma)$  is thus:

$$\frac{1}{2L} \sum_{i=1}^L \lambda(\sigma_i) \mathbb{E}_{p(\mathbf{d}|G)} \mathbb{E}_{q_{\sigma_i}(\tilde{\mathbf{d}}|\mathbf{d}, G)} \left[ \left\| \frac{\mathbf{s}_\theta(\tilde{\mathbf{d}})}{\sigma_i} + \frac{\tilde{\mathbf{d}} - \mathbf{d}}{\sigma_i^2} \right\|_2^2 \right], \quad (8)$$

where  $\lambda(\sigma_i) = \sigma_i^2$  is the coefficient function weighting losses of different noise levels according to Song & Ermon (2019), and all expectations can be efficiently estimated using Monte Carlo estimation.

### 4.3. Conformation Generation via Annealed Langevin Dynamics

Given the molecular graph representation  $G$  and the well-trained noise conditional score networks, molecular confor-

mations are generated using annealed Langevin dynamics, i.e., gradually anneal down the noise level from large to small. We start annealed Langevin dynamics by first sampling an initial conformation  $\mathbf{R}_0$  from some fixed prior distribution, e.g., uniform distribution or Gaussian distribution. Empirically, the choice of the prior distribution is arbitrary as long as the supports of the perturbed distance distributions cover the prior distribution, and we take the Gaussian distribution as prior distribution in our case. Then, we update the conformations by sampling from a series of trained noise conditional score networks  $\{\mathbf{s}_\theta(\mathbf{R}, \sigma_i)\}_{i=1}^L$  sequentially. For each noise level  $\sigma_i$ , starting from the final sample from the last noise level, we run Langevin dynamics for  $T$  steps with a gradually decreasing step size  $\alpha_i = \epsilon \cdot \sigma_i^2 / \sigma_L^2$  for each noise level. In specific, at each sampling step  $t$ , we first obtain the interatomic distances  $\mathbf{d}_{t-1}$  based on the current conformation  $\mathbf{R}_{t-1}$ , and calculate the score function  $\mathbf{s}_\theta(\mathbf{R}_{t-1}, \sigma_i)$  via Eq. 3. The conformation is then updated using the gradient information from the score network. The whole sampling algorithm is presented in Algorithm 1.

#### Algorithm 1 Annealed Langevin dynamics sampling

```

input molecular graph  $G$ , noise levels  $\{\sigma_i\}_{i=1}^L$ , the smallest step size  $\epsilon$ , and the number of sampling steps per noise level  $T$ .
1: Initialize conformation  $\mathbf{R}_0$  from a prior distribution
2: for  $i \leftarrow 1$  to  $L$  do
3:    $\alpha_i \leftarrow \epsilon \cdot \sigma_i^2 / \sigma_L^2$   $\triangleright \alpha_i$  is the step size.
4:   for  $t \leftarrow 1$  to  $T$  do
5:      $\mathbf{d}_{t-1} = g_G(\mathbf{R}_{t-1})$   $\triangleright$  get distances from  $\mathbf{R}_{t-1}$ 
6:      $\mathbf{s}_\theta(\mathbf{R}_{t-1}, \sigma_i) \leftarrow \text{convert}(\mathbf{s}_\theta(\mathbf{d}_{t-1}, \sigma_i))$   $\triangleright$  Eq. 3.
7:     Draw  $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I})$ 
8:      $\mathbf{R}_t \leftarrow \mathbf{R}_{t-1} + \alpha_i \mathbf{s}_\theta(\mathbf{R}_{t-1}, \sigma_i) + \sqrt{2\alpha_i} \mathbf{z}_t$ 
9:   end for
10:   $\mathbf{R}_0 \leftarrow \mathbf{R}_T$ 
11: end for
output Generated conformation  $\mathbf{R}_T$ .
```

#### 4.4. Discussion

**Physical Interpretation of Eq. 3** From a physical perspective,  $s_\theta(\mathbf{d}, \sigma)$  reflects how the distances between connected atoms should change to increase the probability density of the conformation  $\mathbf{R}$ , which can be viewed as a network that predicts the repulsion or attraction forces between atoms. Moreover, as shown in Figure 1,  $s_\theta(\mathbf{R}, \sigma)$  can be viewed as resultant forces acting on each atom computed from Eq. 3, where  $(\mathbf{r}_i - \mathbf{r}_j)/d_{ij}$  is a unit vector indicating the direction of each component force and  $s_\theta(\mathbf{d}, \sigma)_{ij}$  specifies the strength of each component force.

**Handling Stereochemistry** Given a molecular graph, our ConfGF generates all possible stereoisomers due to the stochasticity of the Langevin dynamics. Nevertheless, our model is very general and can be extended to handle stereochemistry by predicting the gradients of torsional angles with respect to coordinates, as torsional angles can be calculated in closed-form from 3D coordinates and are also invariant under rotation and translation. Since the stereochemistry study of conformation generation is beyond the scope of this work, we leave it for future work.

## 5. Experiments

Following existing works on conformation generation, we evaluate the proposed method using the following three standard tasks:

- **Conformation Generation** tests models' capacity to learn distribution of conformations by measuring the quality of generated conformations compared with reference ones (Section 5.1).
- **Distributions Over Distances** evaluates the discrepancy with respect to distance geometry between the generated conformations and the ground-truth conformations (Section 5.2).
- **Property Prediction** is first proposed in Simm & Hernandez-Lobato (2020), which estimates chemical properties for molecular graphs based on a set of sampled conformations (Section 5.3). This can be useful in a variety of applications such as drug discovery, where we want to predict molecular properties from microscopic structures.

Below we describe setups that are shared across the tasks. Additional setups are provided in the task-specific sections.

**Data** Following Xu et al. (2021), we use the GEOM-QM9 and GEOM-Drugs (Axelrod & Gomez-Bombarelli, 2020) datasets for the conformation generation task. We randomly draw 40,000 molecules and select the 5 most likely con-

formations<sup>3</sup> for each molecule, and draw 200 molecules from the remaining data, resulting in a training set with 200,000 conformations and a test set with 22,408 and 14,324 conformations for GEOM-QM9 and GEOM-Drugs respectively. We evaluate the distance modeling task on the ISO17 dataset (Simm & Hernandez-Lobato, 2020). We use the default split of Simm & Hernandez-Lobato (2020), resulting in a training set with 167 molecules and the test set with 30 molecules. For property prediction task, we randomly draw 30 molecules from the GEOM-QM9 dataset with the training molecules excluded.

**Baselines** We compare our approach with four state-of-the-art methods for conformation generation. In specific, **CV-GAE** (Mansimov et al., 2019) is built upon conditional variational graph autoencoder, which directly generates atomic coordinates based on molecular graphs. **GraphDG** (Simm & Hernandez-Lobato, 2020) and **CGCF** (Xu et al., 2021) are distance-based methods, which rely on post-processing algorithms to convert distances to final conformations. The difference of the two methods lies in how the neural networks are designed, i.e., VAE vs. Continuous Flow. **RD-Kit** (Riniker & Landrum, 2015) is a classical Euclidean Distance Geometry-based approach. Results of all baselines across experiments are obtained by running their official codes unless stated.

**Model Configuration** ConfGF is implemented in Pytorch (Paszke et al., 2017). The GINs is implemented with  $N = 4$  layers and the hidden dimension is set as 256 across all modules. For training, we use an exponentially-decayed learning rate starting from 0.001 with a decay rate of 0.95. The model is optimized with Adam (Kingma & Ba, 2014) optimizer on a single Tesla V100 GPU. All hyperparameters related to noise levels as well as annealed Langevin dynamics are selected according to Song & Ermon (2020). See Supplementary material for full details.

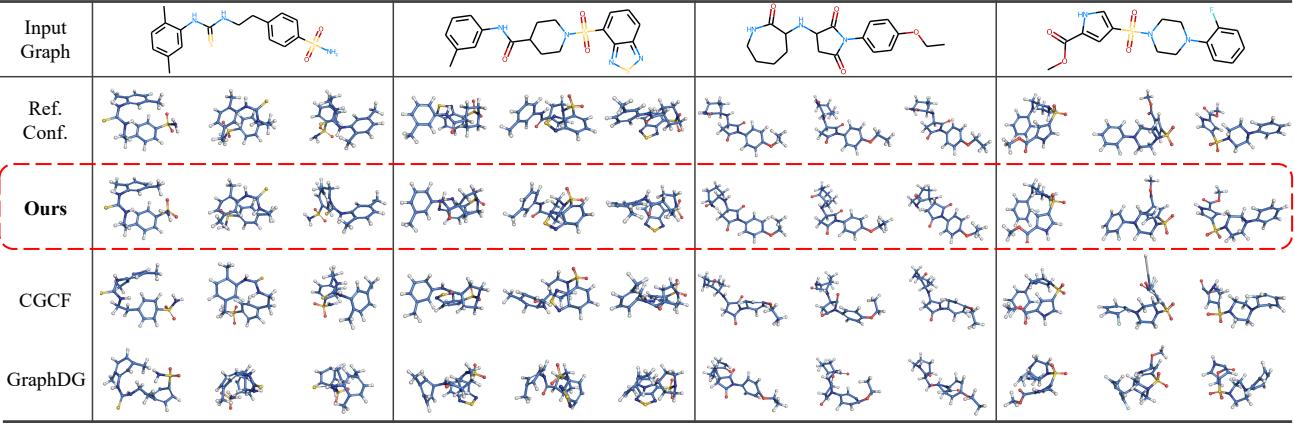
### 5.1. Conformation Generation

**Setup** The first task is to generate conformations with high diversity and accuracy based on molecular graphs. For each molecular graph in the test set, we sample twice as many conformations as the reference ones from each model. To measure the discrepancy between two conformations, following existing work (Hawkins, 2017; Li et al., 2021), we adopt the Root of Mean Squared Deviations (RMSD) of the heavy atoms between aligned conformations:

$$\text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}) = \min_{\Phi} \left( \frac{1}{n} \sum_{i=1}^n \|\Phi(\mathbf{R}_i) - \hat{\mathbf{R}}_i\|^2 \right)^{\frac{1}{2}}, \quad (9)$$

where  $n$  is the number of heavy atoms and  $\Phi$  is an alignment function that aligns two conformations by rotation and trans-

<sup>3</sup>sorted by energy



*Figure 3.* Visualization of conformations generated by different approaches. For each method, we sample multiple conformations and pick ones that are best-aligned with the reference ones, based on four random molecular graphs from the test set of GEOM-Drugs.

lation. Following Xu et al. (2021), we report the **Coverage** (COV) and **Matching** (MAT) score to measure the diversity and accuracy respectively:

$$\text{COV}(S_g, S_r) = \frac{|\{\mathbf{R} \in S_r \mid \text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}) < \delta, \hat{\mathbf{R}} \in S_g\}|}{|S_r|}$$

$$\text{MAT}(S_g, S_r) = \frac{1}{|S_r|} \sum_{\mathbf{R} \in S_r} \min_{\hat{\mathbf{R}} \in S_g} \text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}), \quad (10)$$

where  $S_g$  and  $S_r$  are generated and reference conformation set respectively.  $\delta$  is a given RMSD threshold. In general, a higher COV score indicates a better diversity, while a lower MAT score indicates a better accuracy.

**Results** We evaluate the mean and median COV and MAT scores on both GEOM-QM9 and GEOM-Drugs datasets for all baselines, where mean and median are taken over all molecules in the test set. Results in Table 1 show that our ConfGF achieves the state-of-the-art performance on all four metrics. In particular, as a score-based model that directly estimates the gradients of the log density of atomic coordinates, our approach consistently outperforms other neural models by a large margin. By contrast, the performance of GraphDG and CGCF is inferior to ours due to their two-stage generation procedure, which incurs extra errors when converting distances to conformations. The CVGAE seems to struggle with both of the datasets as it fails to model the roto-translation equivariance of conformations.

When compared with the rule-based method, our approach outperforms the RDKit on 7 out of 8 metrics without relying on any post-processing algorithms, e.g., force fields, as done in previous work (Mansimov et al., 2019; Xu et al., 2021), making it appealing in real-world applications. Figure 3 visualizes several conformations that are best-aligned with the reference ones generated by different methods, which

demonstrates ConfGF’s strong ability to generate realistic and diverse conformations.

*Table 1.* COV and MAT scores of different approaches on GEOM-QM9 and GEOM-Drugs datasets. The threshold  $\delta$  is set as 0.5 Å for QM9 and 1.25 Å for Drugs following Xu et al. (2021). See Supplementary material for additional results.

Dataset	Method	COV (%)		MAT (Å)	
		Mean	Median	Mean	Median
QM9	RDKit	83.26	90.78	0.3447	0.2935
	CVGAE	0.09	0.00	1.6713	1.6088
	GraphDG	73.33	84.21	0.4245	0.3973
	CGCF	78.05	82.48	0.4219	0.3900
	<b>ConfGF</b>	<b>88.49</b>	<b>94.13</b>	<b>0.2673</b>	<b>0.2685</b>
Drugs	RDKit	60.91	65.70	1.2026	<b>1.1252</b>
	CVGAE	0.00	0.00	3.0702	2.9937
	GraphDG	8.27	0.00	1.9722	1.9845
	CGCF	53.96	57.06	1.2487	1.2247
	<b>ConfGF</b>	<b>62.15</b>	<b>70.93</b>	<b>1.1629</b>	1.1596

## 5.2. Distributions Over Distances

**Setup** The second task is to match the generated distributions over distances with the ground-truth. As ConfGF is not designed for modeling molecular distance geometry, we treat interatomic distances calculated from generated conformations as generated distance distributions. For each molecular graph in the test set, we sample 1000 conformations across all methods. Following previous work (Simm & Hernandez-Lobato, 2020; Xu et al., 2021), we report the discrepancy between generated distributions and the ground-truth distributions, by calculating maximum mean discrepancy (MMD) (Gretton et al., 2012) using a Gaussian kernel. In specific, for each graph  $G$  in the test set, we ignore edges connected with the H atoms and evaluate distributions of all distances  $p(d \mid G)$  (All), pairwise distances

**Table 2.** Accuracy of the distributions over distances generated by different approaches compared to the ground-truth. Two different metrics are used: mean and median MMD between ground-truth conformations and generated ones. Results of baselines are taken from Xu et al. (2021).

Method	Single		Pair		All	
	Mean	Median	Mean	Median	Mean	Median
RDKit	3.4513	3.1602	3.8452	3.6287	4.0866	3.7519
CVGAE	4.1789	4.1762	4.9184	5.1856	5.9747	5.9928
GraphDG	0.7645	0.2346	0.8920	0.3287	1.1949	0.5485
CGCF	0.4490	<b>0.1786</b>	0.5509	<b>0.2734</b>	0.8703	0.4447
<b>ConfGF</b>	<b>0.3684</b>	0.2358	<b>0.4582</b>	0.3206	<b>0.6091</b>	<b>0.4240</b>

$p(d_{ij}, d_{uv} \mid G)$  (Pair), and individual distances  $p(d_{ij} \mid G)$  (Single).

**Results** As shown in Table 2, the CVGAE suffers the worst performance due to its poor capacity. Despite the superior performance of RDKit on the conformation generation task, its performance falls far short of other neural models. This is because RDKit incorporates an additional molecular force field (Rappé et al., 1992) to find equilibrium states of conformations after the initial coordinates are generated. By contrast, although not tailored for molecular distance geometry, our ConfGF still achieves competitive results on all evaluation metrics. In particular, the ConfGF outperforms the state-of-the-art method CGCF on 4 out of 6 metrics, and achieves comparable results on the other two metrics.

### 5.3. Property Prediction

**Setup** The third task is to predict the *ensemble property* (Axelrod & Gomez-Bombarelli, 2020) of molecular graphs. Specifically, the ensemble property of a molecular graph is calculated by aggregating the property of different conformations. For each molecular graph in the test set, we first calculate the energy, HOMO and LUMO of all the ground-truth conformations using the quantum chemical calculation package Psi4 (Smith et al., 2020). Then, we calculate ensemble properties including average energy  $\bar{E}$ , lowest energy  $E_{\min}$ , average HOMO-LUMO gap  $\Delta\epsilon$ , minimum gap  $\Delta\epsilon_{\min}$ , and maximum gap  $\Delta\epsilon_{\max}$  based on the conformational properties for each molecular graph. Next, we use ConfGF and other baselines to generate 50 conformations for each molecular graph and calculate the above-mentioned ensemble properties in the same way. We use mean absolute error (MAE) to measure the accuracy of ensemble property prediction. We exclude CVGAE from this analysis due to its poor quality following Simm & Hernandez-Lobato (2020).

**Results** As shown in Table 3, ConfGF outperforms all the machine learning-based methods by a large margin and achieves better accuracy than RDKit in lowest energy  $E_{\min}$  and maximum HOMO-LUMO gap  $\Delta\epsilon_{\max}$ . Closer inspection on the generated samples shows that, although ConfGF has better generation quality in terms of diversity and accu-

racy, it sometimes generates a few outliers which have negative impact on the prediction accuracy. We further present the median of absolute errors in Table 4. When measured by median absolute error, ConfGF has the best accuracy in average energy and the accuracy of other properties is also improved, which reveals the impact of outliers.

**Table 3.** Mean absolute errors (MAE) of predicted ensemble properties. Unit: eV.

Method	$\bar{E}$	$E_{\min}$	$\Delta\epsilon$	$\Delta\epsilon_{\min}$	$\Delta\epsilon_{\max}$
RDKit	<b>0.9233</b>	0.6585	<b>0.3698</b>	<b>0.8021</b>	0.2359
GraphDG	9.1027	0.8882	1.7973	4.1743	0.4776
CGCF	28.9661	2.8410	2.8356	10.6361	0.5954
<b>ConfGF</b>	2.7886	<b>0.1765</b>	0.4688	2.1843	<b>0.1433</b>

**Table 4.** Median of absolute prediction errors. Unit: eV.

Method	$\bar{E}$	$E_{\min}$	$\Delta\epsilon$	$\Delta\epsilon_{\min}$	$\Delta\epsilon_{\max}$
RDKit	0.8914	0.6629	<b>0.2947</b>	<b>0.5196</b>	0.1617
<b>ConfGF</b>	<b>0.5328</b>	<b>0.1145</b>	0.3207	0.7365	<b>0.1337</b>

**Table 5.** Ablation study on the performance of ConfGF. The results of CGCF are taken directly from Table 1. The threshold  $\delta$  is set as 0.5Å for QM9 and 1.25Å for Drugs following Table 1.

Dataset	Method	COV (%)		MAT (Å)	
		Mean	Median	Mean	Median
QM9	CGCF	78.05	82.48	0.4219	0.3900
	ConfGFDist	81.94	85.80	0.3861	0.3571
	<b>ConfGF</b>	<b>88.49</b>	<b>94.13</b>	<b>0.2673</b>	<b>0.2685</b>
Drugs	CGCF	53.96	57.06	1.2487	1.2247
	ConfGFDist	53.40	52.58	1.2493	1.2366
	<b>ConfGF</b>	<b>62.15</b>	<b>70.93</b>	<b>1.1629</b>	<b>1.1596</b>

### 5.4. Ablation Study

So far, we have justified the effectiveness of the proposed ConfGF on multiple tasks. However, it remains unclear whether the proposed strategy, i.e., directly estimating the gradient fields of log density of atomic coordinates, contributes to the superior performance of ConfGF. To gain insights into the working behaviours of ConfGF, we conduct the ablation study in this section.

We adopt a variant of ConfGF, called ConfGFDist, which works in a way similar to distance-based methods (Simm & Hernandez-Lobato, 2020; Xu et al., 2021). In specific, ConfGFDist bypasses chain rule and directly generates interatomic distances via annealed Langevin dynamics using estimated gradient fields of interatomic distances, i.e.,  $s_\theta(\mathbf{d}, \sigma)$  (Section 4.2), after which we convert them to conformations by the same post-processing algorithm as used in Xu et al. (2021). Following the setup of Section 5.1, we evaluate two approaches on the conformation generation task and summarize the results in Table 5. As presented in Table 5, we observe that ConfGFDist performs significantly

worse than ConfGF on both datasets, and performs on par with the state-of-the-art method CGCF. The results indicate that the two-stage generation procedure of distance-based methods does harm the quality of conformations, and our strategy effectively addresses this issue, leading to much more accurate and diverse conformations.

## 6. Conclusion and Future Work

In this paper, we propose a novel principle for molecular conformation generation called ConfGF, where samples are generated using Langevin dynamics within one stage with physically inspired gradient fields of log density of atomic coordinates. We novelly develop an algorithm to effectively estimate these gradients and meanwhile preserve their roto-translation equivariance. Comprehensive experiments over multiple tasks verify that ConfGF outperforms previous state-of-the-art baselines by a significant margin. Our future work will include extending our ConfGF approach to 3D molecular design tasks and many-body particle systems.

## Acknowledgments

We would like to thank Yiqing Jin for refining the figures used in this paper. This project is supported by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, the Canada CIFAR AI Chair Program, collaboration grants between Microsoft Research and Mila, Samsung Electronics Co., Ltd., Amazon Faculty Research Award, Tencent AI Lab Rhino-Bird Gift Fund and a NRC Collaborative R&D Project (AI4D-CORE-06). This project was also partially funded by IVADO Fundamental Research Project grant PRF-2019-3583139727.

## References

- Axelrod, S. and Gomez-Bombarelli, R. Geom: Energy-annotated molecular conformations for property prediction and molecular generation. *arXiv preprint arXiv:2006.05531*, 2020.
- Ballard, A. J., Martiniani, S., Stevenson, J. D., Somani, S., and Wales, D. J. Exploiting the potential energy landscape to sample free energy. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(3):273–289, 2015.
- Blaney, J. and Dixon, J. Distance geometry in molecular modeling. *ChemInform*, 25, 2007.
- Chmiela, S., Sauceda, H. E., Poltavsky, I., Müller, K.-R., and Tkatchenko, A. sgmdl: Constructing accurate and data efficient molecular force fields using machine learning. *Computer Physics Communications*, 240:38–45, 2019.
- De Vivo, M., Masetti, M., Bottegoni, G., and Cavalli, A. Role of molecular dynamics and related methods in drug discovery. *Journal of medicinal chemistry*, 59(9):4035–4061, 2016.
- Duvenaud, D. K., Maclaurin, D., Iparraguirre, J., Bombarell, R., Hirzel, T., Aspuru-Guzik, A., and Adams, R. P. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in neural information processing systems*, pp. 2224–2232, 2015.
- Frenkel, D. and Smit, B. Understanding molecular simulation: from algorithms to applications. 1996.
- Fuchs, F., Worrall, D., Fischer, V., and Welling, M. Se(3)-transformers: 3d roto-translation equivariant attention networks. *NeurIPS*, 2020.
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1263–1272. JMLR.org, 2017.
- Gogineni, T., Xu, Z., Punzalan, E., Jiang, R., Kammeraad, J. A., Tewari, A., and Zimmerman, P. Torsionnet: A reinforcement learning approach to sequential conformer search. *ArXiv*, abs/2006.07078, 2020.
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., and Smola, A. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.
- Halgren, T. A. Merck molecular force field. v. extension of mmff94 using experimental data, additional computational data, and empirical rules. *Journal of Computational Chemistry*, 17(5-6):616–641, 1996.
- Hawkins, P. C. Conformation generation: the state of the art. *Journal of Chemical Information and Modeling*, 57(8):1747–1756, 2017.
- Hu, W., Liu, B., Gomes, J., Zitnik, M., Liang, P., Pande, V., and Leskovec, J. Strategies for pre-training graph neural networks. In *International Conference on Learning Representations*, 2020.
- Hu, W., Shuaibi, M., Das, A., Goyal, S., Sriram, A., Leskovec, J., Parikh, D., and Zitnick, L. Forcenet: A graph neural network for large-scale quantum chemistry simulation. 2021.
- Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. *arXiv preprint arXiv:1802.04364*, 2018.
- Kanal, I. Y., Keith, J. A., and Hutchison, G. A sobering assessment of small-molecule force field methods for low energy conformer predictions. *International Journal of Quantum Chemistry*, 118, 2017.

- Kearnes, S., McCloskey, K., Berndl, M., Pande, V., and Riley, P. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2014.
- Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.
- Klicpera, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. In *International Conference on Learning Representations*, 2020.
- Köhler, J., Klein, L., and Noe, F. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, 2020.
- Li, Z., Yang, S., Song, G., and Cai, L. Conformation-guided molecular representation with hamiltonian neural networks. In *International Conference on Learning Representations*, 2021.
- Liberti, L., Lavor, C., Maculan, N., and Mucherino, A. Euclidean distance geometry and applications. *SIAM review*, 56(1):3–69, 2014.
- Mansimov, E., Mahmood, O., Kang, S., and Cho, K. Molecular geometry prediction using a deep generative graph neural network. *arXiv preprint arXiv:1904.00314*, 2019.
- Meng, C., Yu, L., Song, Y., Song, J., and Ermon, S. Autoregressive score matching. *NeurIPS*, 2020.
- Miller, B., Geiger, M., Smidt, T., and Noé, F. Relevance of rotationally equivariant convolutions for predicting molecular properties. *ArXiv*, abs/2008.08461, 2020.
- Parr, R. Density-functional theory of atoms and molecules. 1989.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. Automatic differentiation in pytorch. In *NIPS-W*, 2017.
- Rappé, A. K., Casewit, C. J., Colwell, K., Goddard III, W. A., and Skiff, W. M. 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.
- 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.
- Satorras, V. G., Hoogeboom, E., and Welling, M. E(n) equivariant graph neural networks, 2021.
- Schütt, K., Kindermans, P.-J., Sauceda Felix, H. E., Chmiela, S., Tkatchenko, A., and Müller, K.-R. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *Advances in Neural Information Processing Systems*, pp. 991–1001. Curran Associates, Inc., 2017.
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R., and Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8: 13890, 2017.
- Shi, C., Xu, M., Zhu, Z., Zhang, W., Zhang, M., and Tang, J. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*, 2020.
- Shim, J. and MacKerell Jr, A. D. Computational ligand-based rational design: role of conformational sampling and force fields in model development. *MedChemComm*, 2(5):356–370, 2011.
- Simm, G. and Hernandez-Lobato, J. M. A generative model for molecular distance geometry. In III, H. D. and Singh, A. (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 8949–8958. PMLR, 2020.
- Simm, G. N. C., Pinsler, R., Csányi, G., and Hernández-Lobato, J. M. Symmetry-aware actor-critic for 3d molecular design. In *International Conference on Learning Representations*, 2021.
- Smith, D. G. A., Burns, L., Simmonett, A., Parrish, R., Schieber, M. C., Galvelis, R., Kraus, P., Kruse, H., Remigio, R. D., Alenaizan, A., James, A. M., Lehtola, S., Misiewicz, J. P., et al. Psi4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of chemical physics*, 2020.
- Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32, pp. 11918–11930. Curran Associates, Inc., 2019.
- Song, Y. and Ermon, S. Improved techniques for training score-based generative models. *NeurIPS*, 2020.
- Song, Y., Garg, S., Shi, J., and Ermon, S. Sliced score matching: A scalable approach to density and score estimation. In Adams, R. P. and Gogate, V. (eds.), *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115, pp. 574–584. PMLR, 2020.

Thomas, N., Smidt, T., Kearnes, S. M., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *ArXiv*, 2018.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Comput.*, 2011.

Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. In *NeurIPS*, 2018.

Welling, M. and Teh, Y. W. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, pp. 681–688, 2011.

Xie, Y., Shi, C., Zhou, H., Yang, Y., Zhang, W., Yu, Y., and Li, L. {MARS}: Markov molecular sampling for multi-objective drug discovery. In *International Conference on Learning Representations*, 2021.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

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.

You, J., Liu, B., Ying, Z., Pande, V., and Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in neural information processing systems*, pp. 6410–6421, 2018.

Zhang, L., Han, J., Wang, H., Car, R., and E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Physical Review Letters*, 120(14):143001, 2018.

## Supplementary Material

### A. Proof of Proposition 1

Let  $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{|V|}) \in \mathbb{R}^{|V| \times 3}$  denote a molecular conformation. Let  $\gamma_t : \mathbb{R}^{|V| \times 3} \rightarrow \mathbb{R}^{|V| \times 3}$  denote any 3D translation function where  $\gamma_t(\mathbf{R})_i := \mathbf{r}_i + \mathbf{t}$ , and let  $\rho_D : \mathbb{R}^{|V| \times 3} \rightarrow \mathbb{R}^{|V| \times 3}$  denote any 3D rotation function whose rotation matrix representation is  $D \in \mathbb{R}^{3 \times 3}$ , i.e.,  $\rho_D(\mathbf{R})_i = D\mathbf{r}_i \in \mathbb{R}^3$ . In our context, we say a score function of atomic coordinates  $\mathbf{s} : \mathbb{R}^{|V| \times 3} \rightarrow \mathbb{R}^{|V| \times 3}$  is roto-translation equivariant if it satisfies:

$$\mathbf{s} \circ \gamma_t \circ \rho_D(\mathbf{R}) = \rho_D \circ \mathbf{s}(\mathbf{R}), \quad (11)$$

Intuitively, the above equation says that applying any translation  $\gamma_t$  to the input has no effect on the output, and applying any rotation  $\rho_D$  to the input has the same effect as applying it to the output, i.e., the gradients rotate together with the molecule system and are invariant under translation.

*Proof.* Denote  $\hat{\mathbf{R}} = \gamma_t \circ \rho_D(\mathbf{R})$ . According to the definition of translation and rotation function, we have  $\hat{\mathbf{R}}_i = \hat{\mathbf{r}}_i = D\mathbf{r}_i + \mathbf{t}$ . According to Eq. 3, we have

$$\begin{aligned} \forall i, (\mathbf{s} \circ \gamma_t \circ \rho_D(\mathbf{R}))_i &= \mathbf{s}(\hat{\mathbf{R}})_i \\ &= \sum_{j \in N(i)} \frac{1}{\hat{d}_{ij}} \cdot \mathbf{s}(\hat{\mathbf{d}})_{ij} \cdot (\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j) \\ &= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \mathbf{s}(\mathbf{d})_{ij} \cdot ((D\mathbf{r}_i + \mathbf{t}) - (D\mathbf{r}_j + \mathbf{t})) \\ &= \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \mathbf{s}(\mathbf{d})_{ij} \cdot D(\mathbf{r}_i - \mathbf{r}_j) \\ &= D \left( \sum_{j \in N(i)} \frac{1}{d_{ij}} \cdot \mathbf{s}(\mathbf{d})_{ij} \cdot (\mathbf{r}_i - \mathbf{r}_j) \right) \\ &= D\mathbf{s}(\mathbf{R})_i \\ &= (\rho_D \circ \mathbf{s}(\mathbf{R}))_i. \end{aligned} \quad (12)$$

Here  $d_{ij} = \hat{d}_{ij}$  and  $\mathbf{d} = \hat{\mathbf{d}}$  because rotation and translation will not change interatomic distances. Together with Eq. 11 and Eq. 12, we conclude that the score network defined in Eq. 3 is roto-translation equivariant.  $\square$

### B. Additional Hyperparameters

We summarize additional hyperparameters of our ConfGF in Table 6, including the biggest noise level  $\sigma_1$ , the smallest noise level  $\sigma_L$ , the number of noise levels  $L$ , the number of sampling steps per noise level  $T$ , the smallest step size  $\epsilon$ , the batch size and the training epochs.

Table 6. Additional hyperparameters of our ConfGF.

Dataset	$\sigma_1$	$\sigma_L$	$L$	$T$	$\epsilon$	Batch size	Training epochs
GEOM-QM9	10	0.01	50	100	2.4e-6	128	200
GEOM-Drugs	10	0.01	50	100	2.4e-6	128	200
ISO17	3	0.1	30	100	2.0e-4	128	100

## C. Additional Experiments

We present more results of the Coverage (COV) score at different threshold  $\delta$  for GEOM-QM9 and GEOM-Drugs datasets in Tables 7 and 8 respectively. Results in Tables 7 and 8 indicate that the proposed ConfGF consistently outperforms previous state-of-the-art baselines, including GraphDG and CGCF. In addition, we also report the Mismatch (MIS) score defined as follows:

$$\text{MIS}(S_g, S_r) = \frac{1}{|S_g|} \left| \left\{ \mathbf{R} \in S_g \mid \text{RMSD}(\mathbf{R}, \hat{\mathbf{R}}) > \delta, \forall \hat{\mathbf{R}} \in S_r \right\} \right|, \quad (13)$$

where  $\delta$  is the threshold. The MIS score measures the fraction of generated conformations that are not matched by any ground-truth conformation in the reference set given a threshold  $\delta$ . A lower MIS score indicates less bad samples and better generation quality. As shown in Tables 7 and 8, the MIS scores of ConfGF are consistently lower than the competing baselines, which demonstrates that our ConfGF generates less invalid conformations compared with the existing models.

Table 7. COV and MIS scores of different approaches on GEOM-QM9 dataset at different threshold  $\delta$ .

QM9			Mean COV (%)			Median COV (%)			Mean MIS (%)			Median MIS (%)			
$\delta$ (Å)	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF
0.10	1.03	0.14	17.59	0.00	0.00	10.29	99.70	99.93	92.30	100.00	100.00	96.23			
0.20	13.05	10.97	43.60	2.99	3.95	37.92	96.12	96.91	81.67	99.04	99.04	85.65			
0.30	32.26	31.02	61.94	18.81	22.94	59.66	87.15	89.21	73.06	94.44	94.44	77.27			
0.40	53.53	53.65	75.45	50.00	52.63	80.64	72.60	78.35	65.38	82.63	82.63	70.00			
0.50	73.33	78.05	88.49	84.21	82.48	94.13	56.09	63.51	53.56	64.66	64.66	56.59			
0.60	88.24	94.85	97.71	98.83	98.79	100.00	40.36	44.82	34.78	43.73	43.73	35.86			
0.70	95.93	99.05	99.52	100.00	100.00	100.00	27.93	29.64	21.00	23.38	23.38	15.64			
0.80	98.70	99.47	99.68	100.00	100.00	100.00	19.15	20.98	12.86	10.72	10.72	5.23			
0.90	99.33	99.50	99.77	100.00	100.00	100.00	12.76	16.74	8.98	3.65	3.65	1.53			
1.00	99.48	99.50	99.86	100.00	100.00	100.00	8.00	14.19	6.76	0.47	0.47	0.36			
1.10	99.51	99.51	99.91	100.00	100.00	100.00	4.99	12.26	5.57	0.00	0.00	0.00			
1.20	99.51	99.51	99.94	100.00	100.00	100.00	2.95	9.68	3.48	0.00	0.00	0.00			
1.30	99.51	99.51	99.96	100.00	100.00	100.00	1.65	7.48	1.91	0.00	0.00	0.00			
1.40	99.51	99.51	99.96	100.00	100.00	100.00	0.84	5.94	1.05	0.00	0.00	0.00			
1.50	99.52	99.51	99.97	100.00	100.00	100.00	0.41	4.94	0.70	0.00	0.00	0.00			

Table 8. COV and MIS scores of different approaches on GEOM-Drugs dataset at different threshold  $\delta$ .

Drugs			Mean COV (%)			Median COV (%)			Mean MIS (%)			Median MIS (%)			
$\delta$ (Å)	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF	GraphDG	CGCF	ConfGF
0.25	0.00	0.06	0.17	0.00	0.00	0.00	100.00	99.99	99.97	100.00	100.00	100.00			
0.50	0.26	0.80	1.15	0.00	0.00	0.00	99.95	99.80	99.52	100.00	100.00	100.00			
0.75	0.75	5.81	9.15	0.00	0.00	0.50	99.69	97.86	96.94	100.00	100.00	99.75			
1.00	2.39	24.67	30.60	0.00	11.81	18.89	99.14	90.82	89.63	100.00	96.50	95.58			
1.25	8.27	53.96	62.15	0.00	57.06	70.93	97.92	78.32	76.58	100.00	86.28	84.48			
1.50	19.96	79.37	86.62	4.00	92.46	98.79	94.40	63.80	60.06	99.14	66.39	63.81			
1.75	36.86	91.47	96.53	26.58	100.00	100.00	87.68	49.72	43.63	95.83	47.09	41.72			
2.00	55.79	96.73	98.62	55.26	100.00	100.00	76.99	37.53	29.80	87.35	30.90	22.44			
2.25	71.43	99.05	99.83	80.00	100.00	100.00	61.76	27.30	18.68	69.74	20.07	10.93			
2.50	83.53	99.47	100.00	95.45	100.00	100.00	44.32	18.97	11.09	42.96	12.33	3.31			
2.75	91.09	99.60	100.00	100.00	100.00	100.00	27.92	12.52	6.32	16.67	6.82	0.74			
3.00	95.00	99.96	100.00	100.00	100.00	100.00	15.97	7.67	3.36	2.46	3.32	0.00			

## D. More Generated Samples

We present more visualizations of generated conformations from our ConfGF in Figure 4, including models trained on GEOM-QM9 and GEOM-drugs datasets.

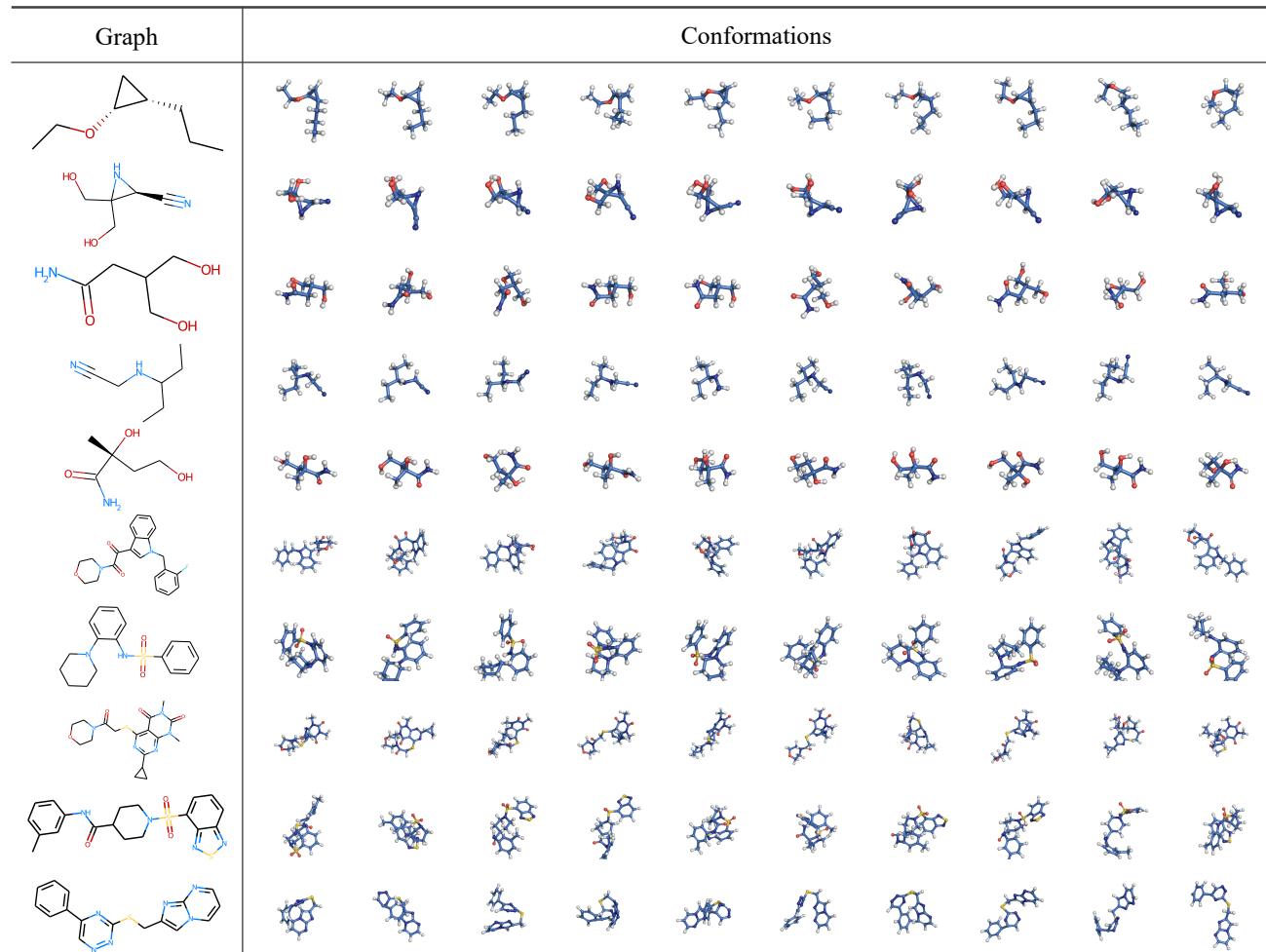


Figure 4. Visualization of conformations generated by our ConfGF. For each molecular graph, we randomly sample 10 generated conformations from our model. For the top 5 rows, the molecular graphs are drawn from the GEOM-QM9 test dataset, while for the bottom 5 rows, the molecular graphs are drawn from the GEOM-drugs test dataset.