EC-Conf: An Ultra-fast Diffusion Model for Molecular Conformation Generation with Equivariant Consistency

Anonymous authors
Paper under double-blind review

Abstract

Despite recent advancement in 3D molecule conformation generation driven by diffusion models, its high computational cost in iterative diffusion/denoising process limits its application. In this paper, an equivariant consistency model (EC-Conf) was proposed as a fast diffusion method for low-energy conformation generation. In EC-Conf, a modified SE (3)-equivariant transformer model was directly used to encode the Cartesian molecular conformations and a highly efficient consistency diffusion process was carried out to generate molecular conformations. It was demonstrated that, with only one sampling step, it can already achieve comparable quality to other diffusion-based models running with thousands denoising steps. Its performance can be further improved with a few more sampling iterations. The performance of EC-Conf is evaluated on both GEOM-QM9 and GEOM-Drugs sets. Our results demonstrate that the efficiency of EC-Conf for learning the distribution of low energy molecular conformation is at least two magnitudes higher than current SOTA diffusion models and could potentially become a useful tool for conformation generation and sampling. We release our code at https://github.com/zhi520/EcConf.

1 Introduction

Three dimensional conformations of a molecule can largely influence its biological and physical properties and biological active conformations are usually low-energy conformers (Perola & Charifson (2004)). Thus, many drug design strategies, including structure-based or ligand-based virtual screening, three-dimensional quantitative structure-activity relationships (QSARs) (Cruciani et al. (2003)), and pharmacophore modeling (Schwab (2010)), require a fast 3D conformation generation and elaboration protocol for sampling biologically relevant conformations. For the former task, programs such as CONCORD (Hendrickson et al. (1993)), CORINA (Gasteiger et al. (1990)), and OMEGA (Hawkins et al. (2010)) are the most popular applications. These approaches make use of known optimal geometries of molecular fragments that are used as templates for constructing reasonable, low energy 3D models of small molecules. These methods usually utilize fallback strategies for structure generation when novel structures appear and belong to rule-based approaches. However, because producing a single 3D structure of a flexible molecule is almost invariably followed by conformational elaboration, it is the latter process that is both the time and quality bottleneck.

With the advancement of deep learning technologies, deep learning methods have been used to modeling the distribution of 3D bioactive conformation and generate 3D conformer directly. The biggest difference between the above-mentioned traditional 3D conformation generators and deep learning based conformational generative models is that those generative models don′t rely on explicit rules to construct 3D conformation but to learn implicitly the distribution of conformational data. Recently, Simm & Hernández-Lobato (2019); Xu et al. (2021c) employed variational autoencoders (VAEs) (Kingma & Welling (2013)) to predict atomic distances and Dinh et al. (2016) made similar effort with flow-based models. Ganea et al. (2021) proposed the GeoMol model, an end-to-end trainable, non-autoregressive conformational generative model consisting of MPNN (Gilmer et al. (2017)) and self-attention layers (Vaswani et al. (2017)) to predict both bond lengths and angles. Application of diffusion model-based technology for conformation generation has gained a lot of attention in the latest two years. Shi et al. (2021) proposed diffusion models by employing denoising score matching

(Song & Ermon (2019; 2020)) to estimate the gradient fields over atomic distances, through which the gradient fields over atomic coordinates can be calculated. They further developed GeoDiff (Xu et al. (2022a)) method, which improves the quality of generated conformations model by directly predicting the coordinates without using intermediate distances.

Although various diffusion-based models were proposed for conformation generation, most of them suffered from large iterative diffusion/denoising steps which makes the inference process too slow. Here, we propose an equivariant consistency model (EC-Conf) for fast diffusion of molecular conformation, largely achieving a balance between the conformation quality and computational efficiency. EC-Conf is inspired from the consistency model (Song et al. (2023)) based on the probability flow with ordinary differential equation, smoothly transform the conformation distribution into a noise distribution with a tractable trajectory satisfied SE (3)-equivariance. Instead of solving the reverse-time SDE in diffusion models, EC-Conf can either directly maps noise vector of prior Gaussian distribution to low-energy conformation in Cartesian coordinates or iteratively maps the solution in the same diffusion trajectory to molecular conformation. These characteristics are critical for greatly decreasing the number of iteration in SDE-based diffusion methods while keep the its high capacity to approximate the complex distribution of conformations. The performance of EC-Conf is evaluated on both GEOM-QM9 and GEOM-Drugs datasets. Our model out-perform non-diffusion models and is comparable with SOTA diffusion models, but with at least two magnitudes higher denoising efficiency.
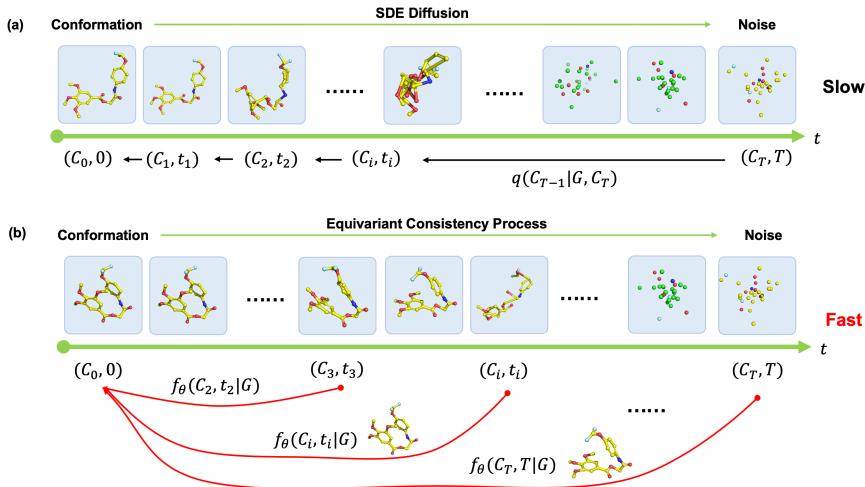

Figure 1: (a) the diffusion process and the generative phase in normal diffusion models. (b) the equivariant consistency model smoothly transforms the conformation data to the noise among a trackable probability flow, and maps any solution on this trajectory to its origin for fast generation.

2 RELATED WORK

2.1 DEEP-LEARNING BASED CONFORMATION GENERATIVE MODELS

With the advancement of deep learning technologies, deep learning methods have been used to directly learn the probability distribution of molecular structures and bring in paradigm change on structure generation. Mansimov et al. (2019) reported the early attempt CVGAE to generate 3D conformations in Cartesian coordinates using the variational autoencoder (VAE) architecture in one-shot. However, its performance is not comparable with traditional rule-based methods. Simm & Hernández-Lobato (2019) proposed a conformation generative model based on distance geometry instead of directly modeling the distributions on cartesian coordinates, named Graph-DG. Dinh et al. (2016) proposed GeoMol, which firstly builds the local structure (LS) by predicting the coordinates of non-terminal atoms, and then refine the LS by the predicted distances and dihedral angles. The quality of generated conformation in these one-shot methods has reached the level of traditional

methods on small molecules, while there are still improving rooms for conformation generation of drug-like molecules.

Another kind of attempts generate and refine the conformation via a set of sequential samplings instead of one-shot generation. Xu et al. (2021a) proposed CGCF combining the advantages of normalizing flows and energy-based approaches to improve the capacity of estimating the multimodal conformation distribution. Then, Xu et al. (2021a) proposed another score-based method ConfGF reported by Shi et al. (2021) learns the pseudo-force on each atom via force matching and obtains new conformations via Langevin Markov chain Monte Carlo (MCMC) sampling on the distance geometry (Shi et al. (2021)). Its performance on the GEOM-Drugs dataset (Axelrod & Gomez-Bombarelli (2022)) is comparable to that of a rule-based method called experimental-torsion-knowledge distance geometry (ETKDG), which is a conformation generation model implemented in RDKit (Landrum (2013)). Scharfer et al. (2013) proposed DMCG trained with a dedicated loss function (Zhu et al. (2022)). GeoDiff (Xu et al. (2022a)) and SDEGen (Zhang et al. (2023)) used a diffusion-based method for conformation generation and also directly predicts the coordinates without using intermediate distances. Similarly, models based on torsion diffusion was also proposed to generate conformations in torsion space instead of cartesian coordinates (Jing et al. (2022)). Although these methods improve the quality of generated conformations of drug-like molecules, the slow sampling speed, which due to large number of diffusion iterations, still limit its applications.

2.2 DIFFUSIONS MODELS

As mentioned above, various deep generative models have been proposed for conformation generation. These methods could be divided into one-shot model or iterative refinement model. Diffusion-based method belongs to the second class.

Diffusion model is a family of probabilistic generative model that progressively destruct data by injecting noise, then learn to reverse this process for data generation. It contains three major categories: denoising diffusion probabilistic models (DDPMs) (Ho et al. (2020)), score-based generative models (SGMs) (Song & Ermon (2019)) and stochastic differential equations (Score SDEs) (Song et al. (2020b)). DDPMs and SGMs can be further generalized to the case of infinite time steps or noise levels where the perturbation and denoising processes are solutions to stochastic differential equations (Yang et al. (2022)). Additionally, Song et al. (2020b) proved the existence of an ordinary differential equation (ODE), namely probability flow ODE (PF ODE), whose trajectories have the same margin as the reverse-time SDE. The main drawback of diffusion model is its demanding of many iterations to solve the SDE or ODE, which is computationally expensive. For example, GeoDiff, a DDPM model, typically requires 5000 iterations for drug-like molecular conformation generation and SDEGen, a Score SDE based method, requires at least 1500 iterations.

Recently, both learning-free sampling method, including critical damped Langevin diffusion (CLD) (Dockhorn et al. (2021)) and denoising diffusion implicit model (DDIM) (Song et al. (2020a)), and learning-based sampling method including optimized discretization (Watson et al. (2021)), knowledge distillation (Salimans & Ho (2022)) and truncated diffusion methods (Lyu et al. (2022)) are proposed to decrease sampling steps. However, how general of these methods for being applied on molecular conformation generation is still unknown. Another type of method is to reformulating the diffusion process. Recently, Xu et al. (2022b) proposed Poisson flow generative models (PFGM), treating the data points as charged particles and then transforming the N-dimensional target-distribution into a uniform angular distribution in $N+1$ dimensional among the Poisson field. It was demonstrated that PFGMs are 10-20 times faster than diffusion models on image generation tasks. More recently, Song et al. (2023) proposed consistency model, a new family of model that generates high quality samples by directly mapping noise to data instead of through the reverse-time SDE. Their results show that consistency models only require a few steps (2-5) of refinement for high quality image generation, which inspired us to incorporate the consistency diffusion process into molecular conformation generation.

3 PRELIMINARIES

3.1 NOTATIONS AND PROBLEM DEFINITION

Notations. For a given molecule, it can be represented with an 2D molecular graph $G = \langle \mathcal{V}, \mathcal{E} \rangle$ including features for all nodes $h_v, \forall v \in \mathcal{V}$, features for all edges $h_{e_{v,w}}, \forall e_{v,w} \in \mathcal{E}$ and a set of coordinates $C = \{c_1, c_2, \ldots c_v\}$ of atoms, $\forall v \in \mathcal{V}$.

Problem Definition. Molecular conformation generation is defined as a conditional generative problem with a given 2D molecular graph G, where we only focus on the low-energy conformations C. For provided multiple molecular graphs $\mathcal{G}$, we aim to learn a transferable approximate generative model $P_\theta(C \mid G)$ mapping the Boltzmann distribution of C under condition of $G \in \mathcal{G}$ to a prior Gaussian distribution, which conformations can be sampled for a given G.

3.2 EQUIVARIANCE

For a molecule conformation, its coordinates are allowed to be changed via translation and rotation in 3D space, while its scalar properties should still be invariant e.g., energy or ADMET properties. This feature is also called SE (3) equivariance, which should be satisfied for 3D based molecular generative models. Formally, a function: $f : x \to y$ is equivariant to a group of transformation $\mathbb{G}$, thus

$$f(D_x(g)x) = D_y(g)f(x), g \in \mathbb{G} \tag{1}$$

where $D_x(g)$ and $D_y(g)$ are transformation matrices parametrized by g in $\mathcal{X}$ and $\mathcal{Y}$. Our EC-Conf models satisfied this SE (3) equivariance to make the estimated likelihood of conformations unaffected by the translation and rotation operation.

4 ECCONF METHOD

4.1 BACKGROUND

Here, we elaborate on the proposed equivariant consistency framework for conformation generation. In DDPM-based diffusion process like GeoDiff, noise from fixed posterior distributions $q\left(C^t \mid C^{t-1}\right)$ is gradually added until the ground truth conformation C^0 is completely destroyed with T time steps. During generation process, an initial state C^T is sampled from standard Gaussian distribution, the conformation is progressively refined via the model learned Markov kernels $p_\theta\left(C^{t-1} \mid G, C^t\right)$ for a given molecular graph G. Song et al. (2020b) have demonstrated that this diffusion process can be described as a discretization process on the time and noise in form of stochastic differential equation (SDE) as defined in equation 2.

$$dC = f(C, t)dt + g(t)dw \tag{2}$$

This process can be reversed by solving the following reverse-time SDE:

$$dC = \left[f(C, t) - g(t)^2 \nabla_C \log p_t(C)\right] dt + g(t)d\bar{w} \tag{3}$$

Where $f(C, t)$ and $g(t)$ are diffusion and drift functions of the SDE, w and $\bar{w}$ are the standard Brownian motion when time flows forward and backwards, respectively, and $\nabla_C \log p_t(C)$ is the gradient of the log probability density. The denoising process can also be described in the form of ordinary differential equation (ODE), representing the probability flow to the same marginals as the reverse-time SDE, i.e. PF ODE (Song et al. (2020b)).

$$dC = \left[f(C, t) - 1/2g(t)^2 \nabla_C \log p_t(C)\right] dt \tag{4}$$

Karras et al. (2022) further simplified the equation 4 by setting $f(C, t) = 0$ and $g(t) = \sqrt{2t}$. In this case, the $p_t(C) = p_{\text{data}}(C) \otimes \mathcal{N}\left(0, T^2 I\right)$ and get the empirical PF ODE as shown in equation 5.

$$\frac{dC_t}{dt} = -t s_\phi(C_t, t) \tag{5}$$

Which allows us to sample $\widehat{C}_T$ from $\pi = \mathcal{N}\left(0, T^2 I\right)$ to initialize the empirical PF ODE and solve it backwards in time with any numerical ODE solver including Euler and Heun solvers. Then we get a solution trajectory $\{C_t\}_{t \in [0, T]}$ and the $\widehat{C}_0$ is an approximate sample from the conformation distribution.

4.2 DEFINITION OF EC-CONF

Given a solution trajectory $\{C_t\}_{t\in[\epsilon,T]}$ of the ODE in equation 5, where ϵ is an approximate to zero. Different from the diffusion model as discussed above, we desire an equivariant consistency function $f : (C_t, t \mid G) \rightarrow C_\epsilon$ which satisfies both a boundary condition and the self-consistency. The boundary condition means that there exists a given function value $f(C_\epsilon, \epsilon) = C_\epsilon$ at the small time step ϵ, while self-consistency means that the output of function $f : (C_t, t \mid G)$ is consistent for arbitrary pairs of (C_t, t) on the same ODE trajectory, i.e., $f(C_t, t \mid G) = f(C_{t'}, t' \mid G)$ for all $t, t' \in [\epsilon, T]$. Once both of these two conditions are satisfied, all the solution $\{C_t\}_{t\in[\epsilon,T]}$ on the ODE trajectory can directly be mapped to the original ground truth C_0. As proved in Song et al. (2023), the transformation between $p_{\text{data}}(C \mid G)$ to $\pi \sim \mathcal{N}(0, T^2 I)$ were built. Additionally, the consistency function should also be SE (3)-equivariant, which means it need to satisfy equation 1. These characteristics not only allow EC-Conf model to make one step generation from the prior distribution but also can improve the quality of generation by chaining the outputs of consistency models at multiple time steps as shown in Figure 1 (b).

We parameterize the consistency model with a learnable function f_θ using skip connections as shown in equations 6, 7 and 8.

$$f_\theta(C, t \mid G) = c_{\text{skip}}(t)C + c_{\text{out}}(t)F_\theta(C, t \mid G) \tag{6}$$

$$c_{\text{skip}}(t) = \frac{\sigma_{\text{data}}^2}{(t-\epsilon)^2 + \sigma_{\text{data}}^2} \tag{7}$$

$$c_{\text{out}}(t) = \frac{\sigma_{\text{data}}(t-\epsilon)}{\sqrt{\sigma_{\text{data}}^2 + t^2}} \tag{8}$$

Where F_θ is a (G, t) related SE (3)-equivariance model to make f_θ satisfied the equivariance. If $c_{\text{skip}}(\epsilon) = 1$ and $c_{\text{out}}(\epsilon) = 0$ for $t = \epsilon$, make $f_\theta(\epsilon) = C_\epsilon$ satisfied the boundary conditions naturally, and both c_{skip}, c_{out}, F_θ are differentiable for training continuous-time consistency model.
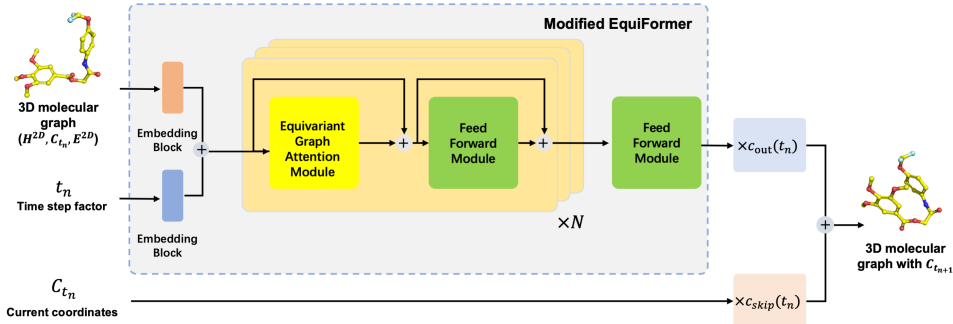

Figure 2: The model architecture in EC-Conf.

4.3 MODEL ARCHITECTURE

In principle, EC-Conf, as a generative learning frame work, could adopt any type of Graph conditioned SE (3)-equivariance neural network for modelling F_θ and is then plugined into the equation 6. Here, we employed and modified the Equiformer Liao & Smidt (2022), a SE (3)-equivariant transformer model based on the irreducible representations and depth-wise tensor production based equivariant attention mechanisms, by incorporating embedding of time step factor, to model the F_θ, and the overall architecture of EC-Conf is shown in Figure 2. The time factor t_n and the atom type are embedded with a linear layer respectively and added together as the input of EquiFormer. As the time factor t_n and atom type are both invariant with coordinate systems, thus, this modification won't affect the SE (3)-Equivariance of Equiformer.

Due to the C_t and F_θ are both SE (3) equivariant, and $c_{\text{skip}}(t)$ and $c_{\text{out}}(t)$ are invariant scalar with given , the SE (3)-equivariance are guaranteed here.

Algorithm 1: Equivariant Consistency Training

input : dataset $\mathcal{D} = \{G_i, C_i\}_{i \in M}$, where M refer to the conformation number in $\mathcal{D}$, initial model parameter θ, learning rate η, step schedule $N(\cdot)$, EMA decay rate schedule $\mu(\cdot)$, $\theta^- \leftarrow \theta$ and $k \leftarrow 0$;

repeat

 Sample $G, C \sim \mathcal{D}$, **and** $n \sim \mathcal{U}[1, N(k) - 1]$;
 Sample $z \sim \mathcal{N}(0, I)$;
 $\mathcal{L}(\theta, \theta^-) \leftarrow MSE(f_\theta(C + t_{n+1} \cdot z, t_{n+1} \mid G), f_{\theta^-}(C + t_n \cdot z, t_n \mid G))$;
 $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}(\theta, \theta^-)$;
 $\theta^- \leftarrow \mu(k)\theta^- + (1 - \mu(k))\theta$;
 $k \leftarrow k + 1$;

until convergence;

4.4 MODEL TRAINING AND CONFORMATION GENERATION

In order to achieve self-consistency in the diffusion model, the training object is to minimize the difference between two adjacent time steps, i.e., $f_\theta(C_{n+1}, t_{n+1} \mid G)$ and $f_\theta(C_n, t_n \mid G)$. As shown in Algorithm 1, for a conformation C sampled from the dataset, we use $C + t_{n+1} \cdot z$ and $C + t_n \cdot z$ to replace C_{n+1}, C_n on the ODE trajectory $\{C_t\}_{t \in [\epsilon, T]}$, thus the f_θ can be trained with the following:

$$\mathcal{L}(\theta) = MSE(f_\theta(C + t_{n+1} \cdot z, t_{n+1} \mid G), f_\theta(C + t_n \cdot z, t_n \mid G)) \tag{9}$$

To make the training process more stable and improve the final performance of f_θ, another function f_{θ^-} with the learnable parameters θ^-, which is the exponential moving average (EMA) of the parameter θ of original function f_θ during training process, is introduced and eventually the difference between $f_\theta(C + t_{n+1} \cdot z, t_{n+1} \mid G)$ and $f_{\theta^-}(C + t_n \cdot z, t_n \mid G)$ is minimized. Here, we refer f_θ as "online network" and the f_{θ^-} as the "target network" as mentioned in Song et al. (2020b) s work. The loss function is reformulated as follows:

$$\mathcal{L}(\theta, \theta^-) = MSE(f_\theta(C + t_{n+1} \cdot z, t_{n+1} \mid G), f_{\theta^-}(C + t_n \cdot z, t_n \mid G)) \tag{10}$$

Parameter θ is updated with stochastic gradient descent, wile θ^- is updated with exponential moving average as shown in equation 11, where μ is the decay rate predefined by EMA schedule.

$$\theta^- = \mu\theta^- + (1 - \mu)\theta \tag{11}$$

Bying doing in this way, we could perform the equivariant consistency training to get the approximate function f_θ to f, namely equivariant consistency conformation generative model (EC-Conf).

In the conformation generation phase, samples are drawn from the initial distribution $\widehat{C}_T \sim N(0, T^2 I)$ and the consistency model is used to generate conformations: $\widehat{C}_\epsilon = f_\theta(C_T, T)$. The conformers are refined with greedy algorithm by alternating denoising and noise injection steps with a set of time points $\{\tau_1, \tau_2 \ldots \tau_{N-1}\}$ as shown in Algorithm 2.

5 EXPERIMENT

This section presents an empirical performance evaluation of EC-Conf on the task of 3D conformation generation for small and drug-like molecules.

5.1 EXPERIMENT SETUP

Following previous work, we also use GEOM-QM9 and GEOM-Drugs datasets for evaluation (Liao & Smidt (2022)). To make a fair benchmark study, we used the same training, validation and test

---

Algorithm 2: Equivariant Consistency Sampling

---

input : Consistency model $f_\theta(\cdot, \cdot)$, sequence of time points $\{\tau_1, \tau_2 \ldots \tau_{N-1}\}$ that

$\tau_1 > \tau_2 > \cdots > \tau_{N-1}$, initial noise $\widehat{C}_T$ sampled from $\mathcal{N}\left(0, T^2 I\right)$, $C \leftarrow f_\theta\left(\widehat{C}_T, T \mid G\right)$;

for $n = 1$ **to** $N - 1$ **do**
    Sample $z \sim \mathcal{N}(0, I)$ **;**
    $\widehat{C}_{\tau_n} \leftarrow C + \sqrt{\tau_n^2 - \epsilon^2} \cdot z$;
    $C \leftarrow f_\theta\left(\widehat{C}_{\tau_n}, \tau_n \mid G\right)$;
end
output: C

---

sets produced by Shi et al. (2021) on both datasets. The GEOM-QM9 is split into training, validation and test set of 39860, 4979, 200 unique molecules, corresponding to 198637, 24825 and 24800 conformations. The GEOM-Drugs set contains 39852, 4983 and 200 molecules in training, validation and test set, corresponding to 198873, 24875 and 14299 conformations, respectively. We examined model performance on the same test set used by GeoDiff model. In current study, following hyperparameters $\sigma_{\text{data}} = 0.5, \epsilon = 10^{-8}, T = 80$ are used. For a given molecule, in case there are K ground truth conformations in test sets, $2K$ conformations are sampled for evaluation.

5.1.1 BASELINES

In our benchmark, 8 recent or established SOTA models including GraphDG (Simm & Hernández-Lobato (2019)), ConfVAE (Xu et al. (2021b)), CGCF (Xu et al. (2021a)), GeoMol (Ganea et al. (2021)), ConfGF Shi et al. (2021), SDE-Gen (Zhang et al. (2023)), GeoDiff (Xu et al. (2022a)) and ETKDG (Landrum (2013)) methods in RDKit. The results of GraphDG, ConfVAE, CGCF, GeoMol, GonfGF, GeoDiff and ETKDG are borrowed from Xu et al. (2022a), while the performance of SDEGen (Zhang et al. (2023)) are evaluated on the same 200 molecules with the provided pretrained models and different settings by ourselves.

5.1.2 EVALUATION METRICS

This task aims to measure the quality and diversity of conformations generated by different models. We follow Ganea et al. (2021) in evaluating four metrics built on the root mean square deviation (RMSD), defined as the normalized Frobenius norm of two atomic coordinate matrices, after alignment by the Kabsch algorithm (Kabsch (1976)). Formally, let S_g and S_r denote the set of generated conformations and the set of reference conformations, respectively, then the coverage and matching measures (Xu et al. (2021a)) following the traditional recall measure can be defined as follows:

$$\text{COV} - \text{R}\left(S_g, S_r\right) = \frac{1}{|S_r|}\left|\left\{\mathcal{C} \in S_r \mid \text{RMSD}(\mathcal{C}, \hat{\mathcal{C}}) \leq \delta, \hat{\mathcal{C}} \in S_g\right\}\right|, \tag{12}$$

$$\text{MAT-R}\left(S_g, S_r\right) = \frac{1}{|S_r|}\sum_{\mathcal{C} \in S_r} \min_{\hat{\mathcal{C}} \in S_g} \text{RMSD}(\mathcal{C}, \hat{\mathcal{C}}) \tag{13}$$

Where δ is a predefined threshold. The other two metrics, COV-P and MAT-P, are inspired by precision and can be defined similarly, but with the generated and referenced set exchanged. In practice, the S_g of each molecule is set to twice the size of the S_r. Intuitively, the COV score measures the percentage of structures in one set that are covered by another set, where coverage means that the RMSD between two constructs is within a certain threshold δ. In contrast, the MAT score measures the average RMSD of one set of conformations against another set of nearest neighbors. In general, a higher COV rate or a lower MAT score indicates that a more realistic conformation is generated. Moreover, the precision metric relies more on quality, while the recall metric focuses more on diversity. Given the specific scenario, either metric may be more attractive. Following previous works (Xu et al. (2021a); Ganea et al. (2021)), δ is set as $0.5Å$ and $1.25Å$ for GEMO-QM9 and GEMO-Drugs datasets respectively.

5.2 RESULTS AND DISCUSSION

5.2.1 THE PERFORMANCE OF EC-CONF WITH DIFFERENT TRAINING TIME STEPS

The first thing to investigate is how training time points influence the model performance. Various models were trained by using different training time steps and evaluated on a random test set. For EC-Conf, user can set the maximal time points in both forward and reverse, corresponding to the training and generation phase. We first evaluated the performance of EC-Conf trained with different maximal time points of 5, 10, 15, 25, 50 in forward ODE process, the iteration steps during generation are the same as training phase and the results are as shown in Table 2 of Appendix A. It seems that both COV-R and MAT-R metrics reached best level when the training time point was set to 25. The performance on COV-R and MAT-R got improved when the training time steps increased from 2 to 25, while it got worse for diffusion step of 50 , which is probably due to the error accumulation during the sampling. For COV-P and MAT-P, the performance of EC-Conf reached the top when training time point was set to 15, and got worse for number of iteration larger than 15. Take both Recall and Precision measurement into consideration, it seems that the models with diffusion steps of 25 in the forward ODE trajectories gave best result in general and it was set as optimum diffusion step in training for the following experiments.

Table 1: The benchmark results on GEOM-Drugs test set

Dataset			Drugs				Drugs			
Task			Conformation Generation				Conformation Generation			
Metric →			COV-R ↑		MAT-R ↓		COV-P ↑		MAT-P ↓	
Model ↓	Steps		Mean	Median	Mean	Median	Mean	Median	Mean	Median
RDKIT	1		0.6091	0.6570	1.2026	1.1252	**0.7222**	**0.8872**	**1.0976**	**0.9539**
GraphDG	1		0.0827	0.0000	1.9722	1.9845	0.0208	0.0000	2.4340	2.4100
ConfVAE	1		0.5520	0.5943	1.2380	1.1417	0.2296	0.1405	1.8287	1.8159
Geomol	1		**0.6716**	**0.7171**	**1.0875**	**1.0586**	—	—	—	—
CGCF	1000		0.5396	0.5706	1.2487	1.2247	0.2168	0.1372	1.8571	1.8066
ConfGF	5000		0.6215	0.7093	1.1629	1.1596	0.2342	0.1552	1.7219	1.6863
SDEGen[a]	1500		0.5601	0.5607	1.2371	1.2303	0.2507	0.1611	1.7196	1.6794
SDEGen[b]	2000		0.5659	0.6358	1.2365	1.2246	0.2644	0.1731	1.7046	1.6997
SDEGen[c]	1500		0.2629	0.1015	1.6011	1.6030	0.1173	0.0302	2.0078	1.9853
SDEGen[d]	6000		0.6727	0.7420	1.1256	1.1289	0.3225	0.2565	1.6793	1.6587
GeoDiff[e]	1000		0.8296	0.9629	0.9525	0.9334	0.4827	0.4603	1.3205	1.2724
GeoDiff-A[f]	5000		0.8836	0.9609	0.8704	0.8628	0.6014	0.6125	1.1864	1.1391
GeoDiff-C[g]	5000		**0.8913**	**0.9788**	**0.8629**	**0.8529**	**0.6147**	**0.6455**	**1.1712**	**1.1232**
EcConf	1		0.7972	0.8419	1.0233	0.9961	0.7049	**0.8370**	1.1183	1.0597
EcConf	2		0.7960	0.8333	1.0157	0.9995	0.7061	0.8297	1.1175	1.0583
EcConf	5		0.8454	0.9118	0.9341	0.9264	0.7140	0.8317	1.0971	1.0270
EcConf	10		0.8576	0.9228	0.9088	0.8975	0.7118	0.8205	1.0925	1.0243
EcConf	15		0.8594	0.9255	0.9046	0.8905	**0.7163**	0.8344	**1.0841**	**1.0176**
EcConf	25		**0.8666**	0.9190	**0.9016**	**0.8869**	0.7136	0.8062	1.0931	1.0307
EcConf	30		0.8640	**0.9344**	0.9024	0.8993	0.7013	0.7945	1.1080	1.0489

[a] SDE-Gen with default sampling settings: $n_{euler} = 250, n_{langevin} = 2, n_{dsm} = 1000$.
[b] SDE-Gen with default sampling settings: $n_{euler} = 500, n_{langevin} = 2, n_{dsm} = 1000$.
[c] SDE-Gen with default sampling settings: $n_{euler} = 500, n_{langevin} = 2, n_{dsm} = 500$.
[d] SDE-Gen with default sampling settings: $n_{euler} = 500, n_{langevin} = 2, n_{dsm} = 5000$.
[e] GeoDiff trains and samples with 1000 steps.
[f] GeoDiff-A trained with alignment approaches.
[g] GeoDiff-A trained with chain-rule approaches.

5.2.2 PERFORMANCE OF EC-CONF UNDER VARIOUS SAMPLING ITERATIONS

Once EC-Conf is trained, it allows either one step generation from the prior Gaussian distribution or carrying out iterative refinement via chaining the outputs of multiple time steps as shown in Algorithm 2. Here, we evaluate the performance of EC-Conf under various sampling iterations. In the meantime, rule based ETKDG method and 7 ML-based baselines were also compared, including one-shot methods: GraphDG, ConfVAE, GeoMol and iterative refinement methods: CGCF, ConfGF, SDEGen, and GeoDiff.

We evaluate our EC-Conf model on QM9 dataset, and the recall measurement of COV-R and MAT-R are as shown in Table 3 of Appendix A, representing the diversity of generated conformations. It's clear that the diffusion-based models out-performed one-shot models suggesting its better capability in reproducing ground truth conformations. Interestingly, one-shot generation of EC-Conf is already comparable with one-shot models, and the result for two iterations is almost the same as SDEGen model running 1500 iterations, i.e., sampling efficiency improved almost 750 times. The performance of the EC-Conf on the diversity gradually converged with more than 5 iterations. Although the recall performance is somewhat lower than GeoDiff model, the improvement on sampling efficiency of 1000 times could justify that EC-Conf model may be a better choice in dealing with large number of molecules. The precision-based metrics of COV-P and MAT-P are also as shown in Table 3 of Appendix A, where the EC-Conf out-performed all the baselines indicating much better quality of conformation generation under all sampling iterations. The optimal performance of EC-Conf was obtained at around 25 iterations.

We also evaluate our EC-Conf model on drug-like molecules with maximum of 50 heavy atoms in GEOM-Drugs sets. The recall-based metrics of COV-R and MAT-R are shown in Table 1. Again, the one-shot generation of EC-Conf greatly out-performed those one-shot baselines and some of the iterative methods such as CGCF, ConfGF, SDEGen. The performance of EC-Conf generation with 5 iterations already performed better than that of GeoDiff with 1000 iterations, representing 200 times better efficiency. The precision-based metrics on GEOM-Drugs set are also listed in Table 1. Although rule based RDKit method generally performed best, performance of EC-Conf model is quite close. Among all deep learning-based models, EC-Conf with single iteration out-performed other models and achieved best results with 15 sampling iterations. Some sample conformations generated by selected models are shown in Figure 3 to provide a qualitative comparison, where EC-Conf is shown to nicely capture both local and global structures in 3D space.
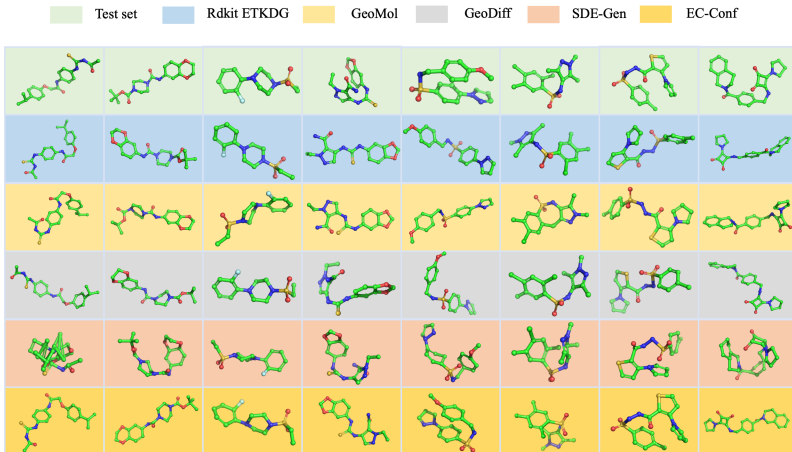

Figure 3: Examples of generated structures for 8 random selected molecules in GEOM-Drugs test set with different methods.

6   CONCLUSION

In this paper, an equivariant consistency model (EC-Conf) was proposed as an ultra-fast diffusion method with only a few iterations for low-energy conformation generation. A time factor-controlled SE (3)-equivariant transformer was used to encode the Cartesian molecular conformations and a highly efficient consistency diffusion process was carried out to generate molecular conformations, largely achieving a balance between the conformation quality and computational efficiency. Our results demonstrate that EC-Conf can potentially learn the distribution of low energy molecular conformation with at least two magnitudes higher efficiency than conventional diffusion models and could potentially become a useful tool for conformation generation and sampling.

REFERENCES

Simon Axelrod and Rafael Gomez-Bombarelli. Geom, energy-annotated molecular conformations for property prediction and molecular generation. *Scientific Data*, 9(1):185, 2022.

Gabriele Cruciani, Emanuele Carosati, and Sergio Clementi. Three-dimensional quantitative structure-property relationships. *The practice of medicinal chemistry*, 2003.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. *arXiv preprint arXiv:2112.07068*, 2021.

Octavian Ganea, Lagnajit Pattanaik, Connor Coley, Regina Barzilay, Klavs Jensen, William Green, and Tommi Jaakkola. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. *Advances in Neural Information Processing Systems*, 34:13757–13769, 2021.

J Gasteiger, C Rudolph, and J Sadowski. Automatic generation of 3d-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology*, 3(6):537–547, 1990.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

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

Marissa A Hendrickson, Marc C Nicklaus, George WA Milne, and D Zaharevitz. Concord and cambridge: comparison of computer generated chemical structures with x-ray crystallographic data. *Journal of chemical information and computer sciences*, 33(1):155–163, 1993.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in Neural Information Processing Systems*, 35:24240–24253, 2022.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32 (5):922–923, 1976.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Greg Landrum. Rdkit documentation. *Release*, 1(1-79):4, 2013.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Zhaoyang Lyu, Xudong Xu, Ceyuan Yang, Dahua Lin, and Bo Dai. Accelerating diffusion models via early stop of the diffusion process. *arXiv preprint arXiv:2205.12524*, 2022.

Elman Mansimov, Omar Mahmood, Seokho Kang, and Kyunghyun Cho. Molecular geometry prediction using a deep generative graph neural network. *Scientific reports*, 9(1):20381, 2019.

Emanuele Perola and Paul S Charifson. Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. *Journal of medicinal chemistry*, 47(10):2499–2510, 2004.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.

Christin Scharfer, Tanja Schulz-Gasch, Hans-Christian Ehrlich, Wolfgang Guba, Matthias Rarey, and Martin Stahl. Torsion angle preferences in druglike chemical space: a comprehensive guide. *Journal of Medicinal Chemistry*, 56(5):2016–2028, 2013.

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

Chence Shi, Shitong Luo, Minkai Xu, and Jian Tang. Learning gradient fields for molecular conformation generation. In *International conference on machine learning*, pp. 9558–9568. PMLR, 2021.

Gregor NC Simm and José Miguel Hernández-Lobato. A generative model for molecular distance geometry. *arXiv preprint arXiv:1909.11459*, 2019.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. *Advances in neural information processing systems*, 33:12438–12448, 2020.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. 2023.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Daniel Watson, Jonathan Ho, Mohammad Norouzi, and William Chan. Learning to efficiently sample from diffusion probabilistic models. *arXiv preprint arXiv:2106.03802*, 2021.

Minkai Xu, Shitong Luo, Yoshua Bengio, Jian Peng, and Jian Tang. Learning neural generative dynamics for molecular conformation generation. *arXiv preprint arXiv:2102.10240*, 2021a.

Minkai Xu, Wujie Wang, Shitong Luo, Chence Shi, Yoshua Bengio, Rafael Gómez-Bombarelli, and Jian Tang. An end-to-end framework for molecular conformation generation via bilevel programming. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 11537–11547. PMLR, 2021b. URL http://proceedings.mlr.press/v139/xu21f.html.

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

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022a.

Yilun Xu, Ziming Liu, Max Tegmark, and Tommi Jaakkola. Poisson flow generative models. *Advances in Neural Information Processing Systems*, 35:16782–16795, 2022b.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Yingxia Shao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *arXiv preprint arXiv:2209.00796*, 2022.

Haotian Zhang, Shengming Li, Jintu Zhang, Zhe Wang, Jike Wang, Dejun Jiang, Zhiwen Bian, Yixue Zhang, Yafeng Deng, Jianfei Song, et al. Sdegen: learning to evolve molecular conformations from thermodynamic noise for conformation generation. *Chemical Science*, 14(6): 1557–1568, 2023.

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

A APPENDIX

Table 2: Results of EC-Conf with different iteration steps on the GEMO-QM9 dataset

	QM9				QM9			
	Conformation Generation				**Conformation Generation**			
Steps	COV-R ↑		MAT-R ↓		COV-P ↑		MAT-P ↓	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
5	0.4585	0.4298	0.5215	0.5202	0.8569	0.9228	0.3555	0.3513
10	0.7586	0.7977	0.3605	0.3521	0.8682	0.9090	0.3396	0.3368
15	0.7628	0.7907	0.3418	0.3383	**0.8945**	**0.9396**	**0.3098**	**0.3045**
25	**0.8235**	**0.8654**	**0.3223**	**0.3196**	0.8630	0.9088	0.3368	0.3356
50	0.8048	0.8571	0.3385	0.3307	0.8279	0.8712	0.3540	0.3513

Table 3: The benchmark results on GEMO-QM9 test set

Dataset			QM9				QM9			
Task			**Conformation Generation**				**Conformation Generation**			
Metric → Model ↓	Steps		COV-R ↑		MAT-R ↓		COV-P ↑		MAT-P ↓	
			Mean	Median	Mean	Median	Mean	Median	Mean	Median
RDKIT	1		**0.8326**	**0.9078**	**0.3447**	**0.2935**	—	—	—	—
GraphDG	1		0.7333	0.8421	0.4245	0.3973	**0.4390**	**0.3533**	**0.5809**	**0.5823**
ConfVAE	1		0.7784	0. 8820	0.4154	0.3739	0.3802	0.3467	0.6215	0.6091
Geomol	1		0.7126	0.7200	0.3731	0.3731	—	—	—	—
CGCF	1000		0.7805	0.8248	0.4219	0.3900	0.3649	0.3357	0.6615	0.6427
ConfGF	5000		0.8849	0.9431	0.2673	0.2685	0.4643	0.4341	0.5224	0.5124
SDEGen[a]	1500		0.8153	0.8599	0.3568	0.3612	0.4837	0.4663	0.5662	0.5483
GeoDiff-A[b]	5000		**0.9054**	**0.9461**	0.2104	0.2021	0.5235	0.5010	0.4539	0.4399
GeoDiff-C[c]	5000		0.9007	0.9339	**0.2090**	**0.1988**	**0.5279**	**0.5029**	**0.4448**	**0.4267**
EcConf	1		0.7541	0.7935	0.4110	0.4085	0.7977	0.8252	0.4137	0.4148
EcConf	2		0.7589	0.7971	0.4087	0.4016	0.7985	0.8285	0.4128	0.4176
EcConf	5		0.8295	0.8806	0.3475	0.3440	0.8324	0.8766	0.3777	0.3732
EcConf	10		**0.8314**	**0.8811**	0.3308	0.3307	0.8507	0.8929	0.3570	0.3559
EcConf	15		0.8298	0.8781	0.3253	0.3244	0.8579	0.9040	0.3476	0.3468
EcConf	25		0.8235	0.8654	**0.3223**	**0.3196**	**0.8630**	**0.9088**	0.3368	0.3356
EcConf	30		0.8131	0.8662	0.3241	0.3223	0.8614	0.9007	**0.3330**	**0.3332**

[a] SDE-Gen with default sampling settings: $n_{\text{euler}} = 250, n_{\text{langevin}} = 2, n_{\text{dsm}} = 1000$.

[b] GeoDiff-A trained with alignment approaches.

[c] GeoDiff-A trained with chain-rule approaches.