EC-Conf: A ultra-fast diffusion model for molecular conformation generation with equivariant consistency

Zhiguang Fan[1,2], Yuedong Yang[1], Mingyuan Xu[2*] and Hongming Chen[2,3*]

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. Here, 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.

Scientific Contributions

In this work, we proposed an equivariant consistency model that significantly improves the efficiency of conformation generation in diffusion-based models while maintaining high structural quality. This method serves as a general framework and can be further extended to more complex structure generation and prediction tasks, including those involving proteins, in future steps.

Key points

- A novel ultra-fast equivariant diffusion model, EC-Conf, was proposed for low-energy conformation generation by construction of a consistency process.
- Compared with other SOTA diffusion models running with thousands denoising steps, EC-Conf can achieve comparable quality with only one sampling step and keep improving with a few more sampling iterations.
- The efficiency of EC-Conf is at least two magnitudes higher than current SOTA diffusion models.
- The EC-Conf is universal and can be easily extended to various conformation generation tasks such as protein–ligand docking pose.

Keywords Conformation generation, Ultra-fast diffusion models

*Correspondence:
Mingyuan Xu
mingyuan.xu.sci@gmail.com
Hongming Chen
chen_hongming@gzlab.ac.cn
Full list of author information is available at the end of the article

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 2 of 15

Introduction

Three dimensional conformations of a molecule can largely influence its biological and physical properties and biological active conformations are usually low-energy conformers [1]. Thus, many drug design strategies, including structure-based [2] or ligand-based virtual screening [3–5], three-dimensional quantitative structure–activity relationships (QSARs) [6], and pharmacophore modeling [7], require a fast 3D conformation generation and elaboration protocol for sampling biologically relevant conformations. For the former task, programs such as CONCORD [8], CORINA [9], and OMEGA [10] 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 learn the distribution of 3D bioactive conformation and generate 3D conformer directly. For a given molecules, one of the most valuable applications of conformation generative model is to generate the conformation ensembles satisfied Boltzmann distributions, which is critical for fast-estimating the free-energy. The biggest difference between the above-mentioned traditional 3D conformation generators and deep learning-based conformation generative models is that those generative models don't rely on explicit rules to construct 3D conformation but learn implicitly the distribution of conformation data. Mansimov and coworkers reported the early attempt CVGAE to generate 3D conformations in Cartesian coordinates using the variational autoencoder (VAE) architecture in one-shot [11]. However, its performance is not comparable with traditional rule-based methods. Simm et al. proposed a conformation generative model based on distance geometry instead of directly modeling the distributions on Cartesian coordinates, named Graph-DG [12], while ConfVAE take the distribution of distances as intermediate variables to generate conformations [13]. Ganea et al. proposed GeoMol [14], which, at the first step, constructs the local structure (LS) by predicting the coordinates of non-terminal atoms, and then refines 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 fragments, while there are still improving rooms for drug-like molecules.

Another type of attempt explores the construction of conformation via a set of sequential samplings instead of one-shot generation. Xu et al. proposed CGCF method combining the advantages of normalizing flows and energy-based approaches to improve the prediction of multimodal conformation distribution [15]. Then, Xu et al. proposed another score-based method ConfGF by learning the pseudo-force on each atom via force matching and obtaining new conformations via Langevin Markov chain Monte Carlo (MCMC) sampling on the distance geometry [16]. Its performance on the GEOM-Drugs dataset 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 [17, 18]. Xia and Liu et al. proposed DMCG trained with a dedicated loss function [19]. GeoDiff [20] and SDEGen [21] employed diffusion model for conformation generation and can directly predict 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 [22]. Although these diffusion methods based on DDPM [23], SGMs [24, 25] or stochastic differential equations (Score SDEs) [26] improve the quality of generated conformations of drug-like molecules, the slow sampling speed, which due to large number of diffusion/denoising iteration, still limit their application.

More recently, Song et al. proposed consistency model, a new diffusion model which can generate high quality samples by directly mapping noise to data instead of through the reverse-time SDE [27]. Their results show that consistency models only require a few steps ($2 \sim 5$) of refinement for high quality image generation, which inspired us to incorporate the consistency diffusion process into molecular conformation generation. Here, we proposed 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 based on the probability flow with ordinary differential equation (ODE) and can smoothly transform the conformation distribution into a noise distribution with a tractable trajectory satisfied SE (3)-equivariance. Instead of solving the reverse-time stochastic differential equation (SDE) in diffusion models, EC-Conf can either directly map noise vector of prior Gaussian distribution to low-energy conformation in Cartesian coordinates or the solutions in the same diffusion trajectory to molecular conformation. These characteristics are

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 3 of 15

critical for decreasing the number of iterations while still keeping its high capacity to approximate the complex distribution of conformations. The performance of EC-Conf is evaluated in both GEOM-QM9 and GEOM-Drugs datasets. Our model out-performs non-diffusion models and is comparable with SOTA diffusion models, but with at least two magnitudes higher denoising efficiency.

Theory and methods

Problem definition

Molecular conformer generation is defined as a conditional generative problem, which generating low-energy conformations C with a given 2D molecular graph G. For given multiple molecular graphs $\mathcal{G}$, we aim to learn a transferable generative model $p_\theta(C|G)$ mapping the Boltzmann distribution of C under condition of $G \in \mathcal{G}$ to a prior Gaussian distribution.

Diffusion process for conformation generations

As mentioned above, high quality conformation generations are still challenging for drug-like molecules and big bio-molecules. Among commonly used generative models, GAN models are suffering from the unstable training and less diversity in generation due to their adversarial training nature [28]. VAE models rely on a surrogate loss, limits its performance on high quality generation [29]. Similarly, Flow-based models have to employ a specialized architectures to construct reversible transform [30]. Compared with GAN, VAE and Flow-based models, diffusion-based methods are generally more stable, robust for training, can capture the underlying data distribution more comprehensively and produce high-quality, high-fidelity conformations but suffering from the sampling efficiency in conformation generations [31].

In DDPM-based diffusion process like GeoDiff, noise from fixed posterior distributions $q(C^t|C^{t-1})$ is gradually added until the ground truth conformation C^0 is completely destroyed with T 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 $p_\theta(C^{t-1}|G, C^t)$ for a given molecular graph G. Song et al. have demonstrated that this diffusion process can be described as a discretization process on the time and noise in form of SDE as defined in Eq. (2). [32]

$$dC = f(C,t)dt + g(t)dw \qquad (1)$$

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

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

where, $f(C,t)$ and $g(t)$ are diffusion and drift functions of the SDE, w and $\overline{w}$ are the standard Brownian motion when time flows forward and backwards respectively, and $\nabla_C log q_t(C)$ is the gradient of the log probability density $\nabla_C log p(C)$. The rough propagation path for between target distributions and prior gaussian distributions largely limits the efficiency of sampling.

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. the PF ODE method:

$$dC = [f(C,t) - 1/2g(t)^2 \nabla_C \log p_t(C)]dt \qquad (3)$$

It allows a smoother propagation between distributions to accelerate the sampling phase. However, ODE trajectory still follows the chain rules, limiting the efficiency.

Karras et al. further simplified the Eq. (3) by setting $f(C,t) = 0$ and $g(t) = \sqrt{2t}$. In this case, the $p_t(C) = p_{data}(C) \otimes \mathcal{N}(0, T^2I)$ and a empirical PF ODE was obtained as shown in Eq. (4) [33]:

$$\frac{dC_t}{dt} = -ts_\phi(C_t, t) \qquad (4)$$

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

Based on the ODE trajectory $\{C_t\}_{t \in [0,T]}$, Song et al. proposed the consistency model by directly mapping the points on the same trajectory to the same initial point in diffusions, greatly improving the sample efficiency [27].

Equivariant consistency models for conformation generations

Inspired from the consistency model, we proposed a novel equivariant diffusion model, named EC-Conf, for ultra-fast diffusion of molecular conformation in cartesian system, largely achieving a balance between the conformation quality and computational efficiency.

Basic Concepts. Given a solution trajectory $\{C_t\}_{t \in [\epsilon, T]}$ of the ODE in Eq. (4), where $\epsilon \to 0$, we desire an equivariant consistency function $f : (C_t, t|G) \to C_\epsilon$ which satisfies both the boundary condition and the self-consistency as shown in Eqs. (5) and (6), respectively.

$$f(C_\epsilon, \epsilon) = C_\epsilon \qquad (5)$$

$$f(C_t, t|G) = f(C_{t'}, t'|G) \forall t, t' \in [\epsilon, T] \qquad (6)$$

As proved by Song et al., all the solutions in $\{C_t\}_{t\in[\epsilon,T]}$ on the ODE trajectory can directly be mapped to the original ground truth C_0 if both of these two conditions are satisfied. It allows a smoother transformation between $p_{data}(C|G)$ to $\pi \sim \mathcal{N}(0, T^2 I)$ to accelerate the diffusion process than DDPM and Score-SDE models.

Additionally, the consistency function should also be SE (3)-equivariant, which means a molecular coordinate C are allowed to be changed via translation and rotation in 3D space, while its scalar properties should still be invariant. Formally, a function $f : \mathcal{X} \to \mathcal{Y}$ is equivariant to a group of transformation $\mathbb{G}$, thus:

$$f\left(D_{\mathcal{X}}(g)C|G\right) = D_{\mathcal{Y}}(g)f(C|G), g \in \mathbb{G} \qquad (7)$$

where $D_{\mathcal{X}}(g)$ and $D_{\mathcal{Y}}(g)$ are transformation matrices parameterized by g in coordinates system $\mathcal{X}$ and $\mathcal{Y}$.

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 Fig. 1b.

Here, we parameterize the equivariant consistency model with a learnable function f_θ using skip connections as shown in Eqs. (8 ~ 10):

$$f_\theta(C,t|G) = c_{skip}(t)C + c_{out}(t)F_\theta(C,t|G) \qquad (8)$$

$$c_{skip}(t) = \frac{\sigma_{data}^2}{(t-\epsilon)^2 + \sigma_{data}^2} \qquad (9)$$

$$c_{out}(t) = \frac{\sigma_{data}(t-\epsilon)}{\sqrt{\sigma_{data}^2 + t^2}} \qquad (10)$$

where F_θ is an (G, t) related noise model fulfilling the SE (3) equivariance. It ensures the SE (3) equivariance of f_θ. It is evident that as $\epsilon \to 0$, $c_{skip}(\epsilon) \to 1$ and $c_{out}(\epsilon) \to 0$ for $t = \epsilon$, making $f_\theta(C, \epsilon|G) = C_\epsilon$, thus satisfying the boundary conditions. Due to $c_{skip}, c_{out},$ and F_θ are both differentiable functions, we can train the f_θ by minimizing the prediction difference between time steps t and t' to satisfy the self-consistency, as illustrated in Eq. (11).

$$L = MSE(f_\theta(C_t, t|G), f_\theta(C_{t'}, t'|G)) \qquad (11)$$

Model Architecture In principle, EC-Conf could utilize any type of graph conditioned equivariant neural network for modelling F_θ. Here, we employed a customized EquiFormer [34] neural network, which is a transformer model based on the irreducible representation and depth-wise tensor production (DTP) based equivariant attention mechanism. This model incorporates embedding of time step factor t_n. The overall architecture of EC-Conf is shown in Fig. 2a. The time factor t_n and the atom
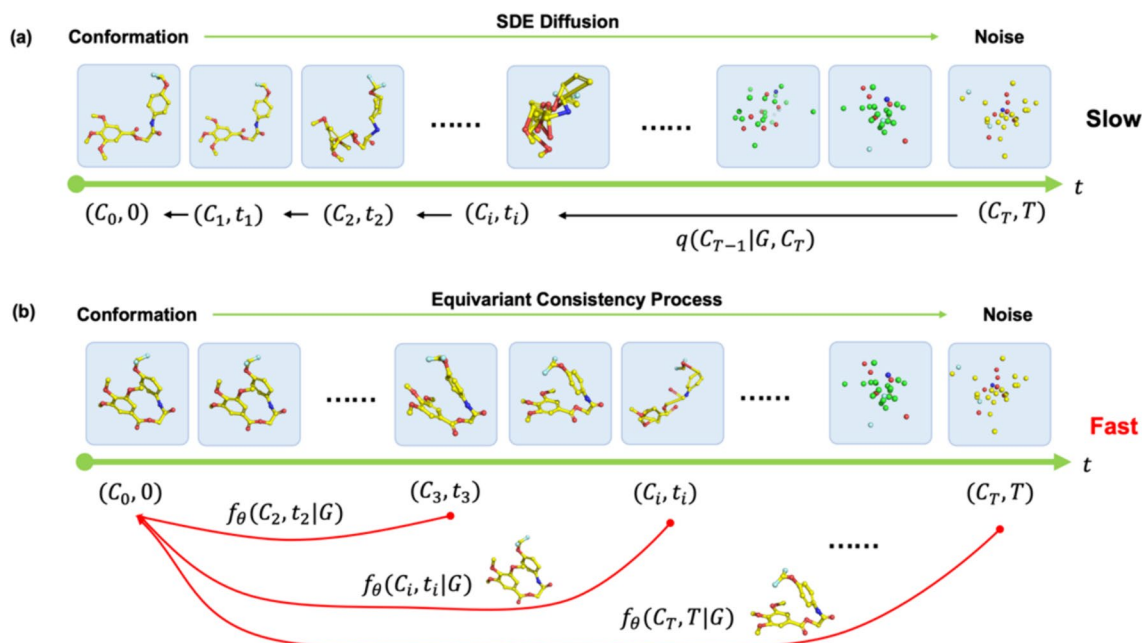
Fig. 1 a The diffusion process and the generative phase in normal diffusion models. **b** The equivariant consistency model smoothly transforms the conformation data to the noise among a tractable probability flow (equivariant consistency process), and maps any solution on this trajectory to its origin for fast generation
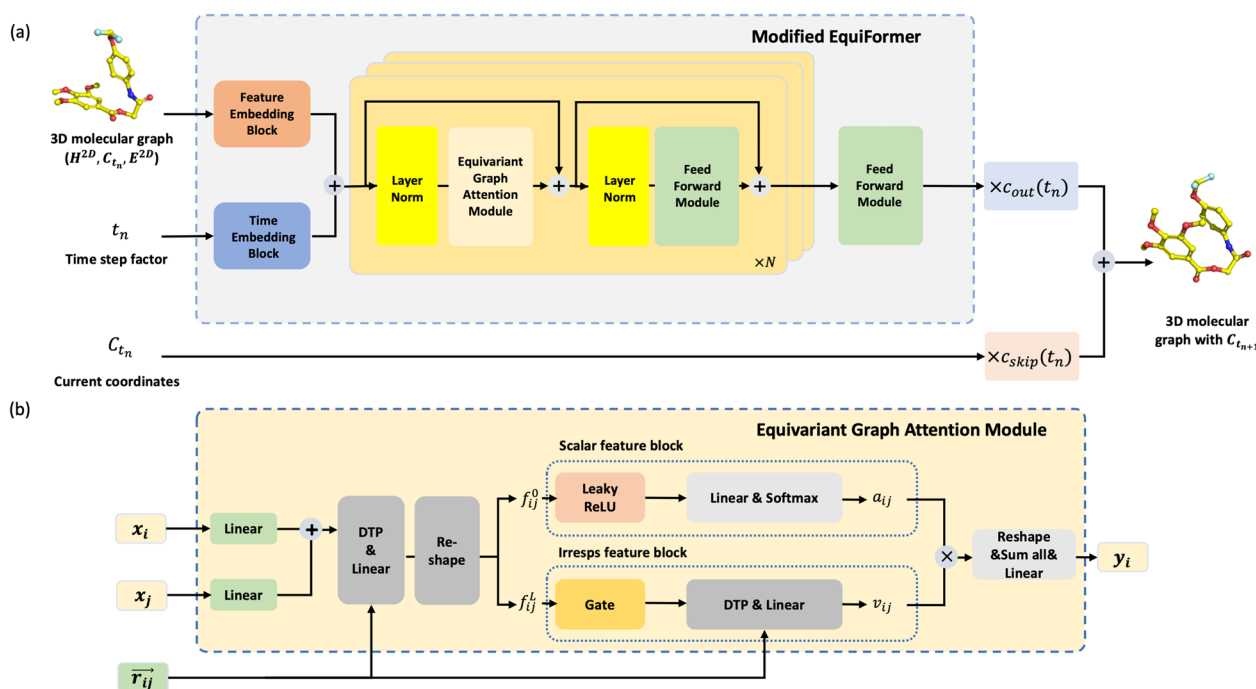
Fig. 2 **a** The model architecture in EC-Conf. **b** The mechanism of equivariant graph attention

type Z_i are embedded with a linear layer respectively and added together as the input of EquiFormer. By repeated stacking equivariant graph attention (EGA) modules and feed-forward modules, Equiformer integrates the content and geometric information to predict the coordinates changes δC_{t_n} from time step t_n. The coordinates $C_{t_{n+1}}$ are then computed by combining C_{t_n} and δC_{t_n} using skip connections.

The architecture of EGA module is illustrated in Fig. 2b. For a pair of neighboring nodes $\{i, j\}$ in graph G, their embeddings x_i, x_j are added after a linear transformation and updated through an equivariant DTP, incorporating their distance r_{ij}. The updated embeddings are reshaped into scalar features f_{ij}^0 and irreps features f_{ij}^L, which are sent to scalar and irreps feature blocks, respectively. Finally, the updated scalar features a_{ij} and irreps features v_{ij} are multiplicated to update x_i. By summing over all neighbors of node i as message passing neural network, both the content and geometric information are incorporated into its embedding x_i. A more detailed description of EGA can be found from reference 34.

As the time factor t_n and atom type are independent to the coordinates, thus, this modification won't affect the SE (3)-equivariance of EquiFormer.

Model Training and Conformer Generation 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 loss function:

$$\mathcal{L}(\theta) = MSE(f_\theta(C + t_{n+1} \cdot z, t_{n+1}|G), f_\theta(C + t_n, t_n|G)) \tag{12}$$

To make the training process more stable and improve the final performance of f_θ, exponential moving average (EMA) technique are adopted as mentioned in Ref 27. Here, we create another function f_{θ^-} with the EMA of original parameter θ during training process, and minimize the difference between $f_\theta(C + t_{n+1} \cdot z, t_{n+1}|G)$ and $f_{\theta^-}(C + t_n \cdot z, t_n|G)$ as Song et al's work [27]. Here, we refer f_θ as "online network" and f_{θ^-} as the "target network" for clarity. The loss function is reformulated as following:

$$\mathcal{L}(\theta, \theta^-) = MSE(f_\theta(C + t_{n+1} \cdot z, t_{n+1}|G), f_{\theta^-}(C + t_n \cdot z, t_n|G)) \tag{13}$$

Parameter θ is updated with stochastic gradient descent, wile θ^- is updated with exponential moving average as shown in Eq. (14), where μ is the decay rate predefined by EMA schedule.

$$\theta^- = \mu\theta^- + (1 - \mu)\theta \tag{14}$$

By doing in this way, we could perform the equivariant consistency training to get the approximate function f_θ to f as shown in Fig. 3a and Algorithm S1.

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 6 of 15
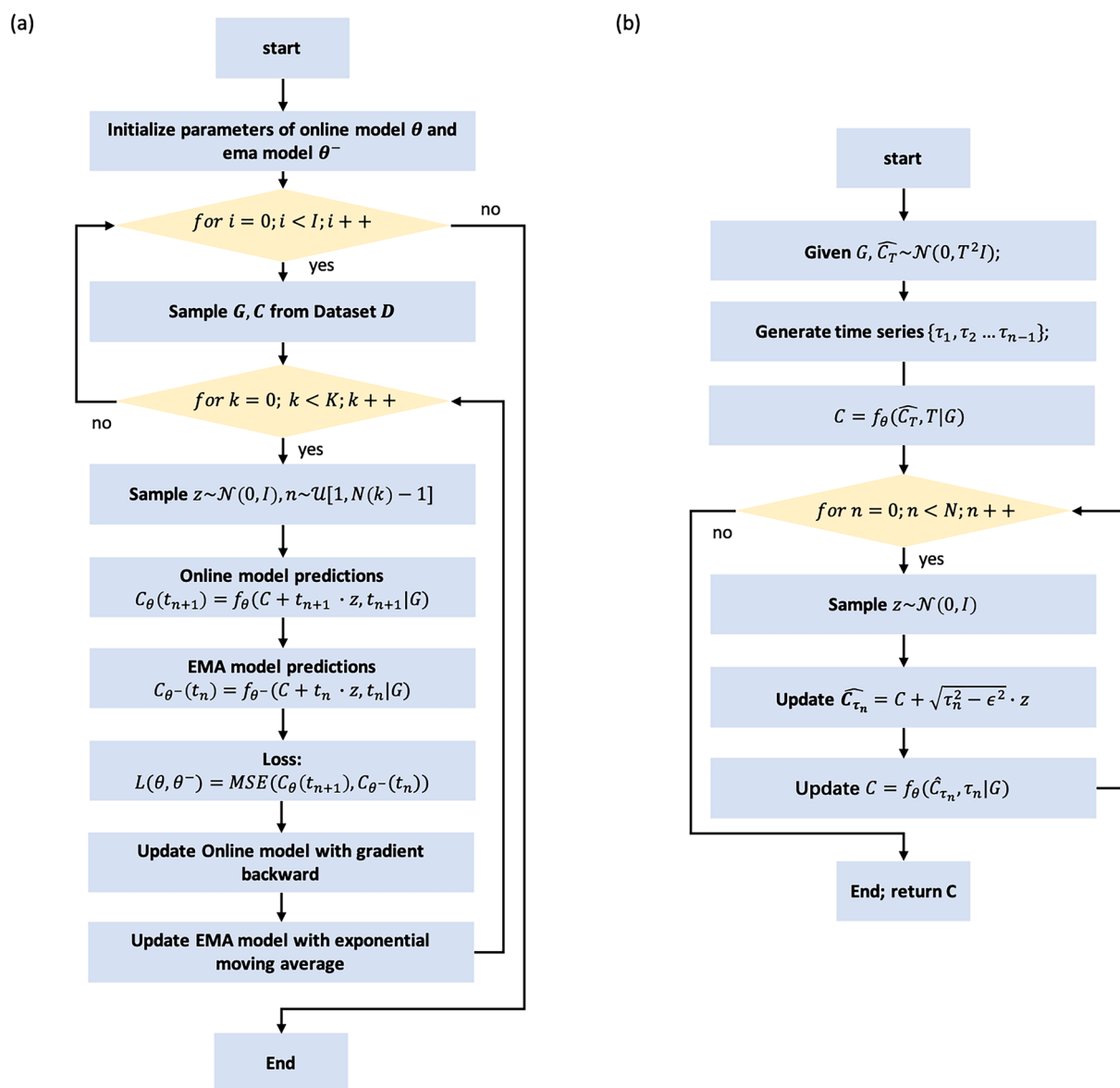
Fig. 3 The training (**a**) and sampling (**b**) workflow of EC-Conf

In the conformer generation phase, samples are drawn from the initial distribution $\widehat{C}_T \sim N(0, T^2I)$ and the consistency model is used to generate conformers: $\widehat{C}_\epsilon = f_\theta(C_T, T)$. The conformers are refined with greedy algorithm by alternating denoising and noise injection steps with a set of time points $\{\tau_1, \tau_2 \cdots \tau_{N-1}\}$ as shown in Fig. 3b and Algorithm S2.

Computational details
Following previous work, we also used GEOM-QM9 [35] and GEOM-Drugs datasets [36] for evaluation. To make a fair benchmark study, we used the same training, validation and test sets produced by Shi et al. for both datasets [13]. Specifically, the GEOM-QM9 is split into training, validation and test set of 39,860, 4979, 200 unique molecules, corresponding to 199,300, 24,895 and 797 conformations; The GEOM-Drugs set contains 39,852, 4983 and 200 molecules in training, validation and test set, corresponding to 199,260, 24,915 and 15,864 conformations, respectively. We examined model performance on the same test set used by GeoDiff model. [20] In current study, following hyperparameters $\sigma_{data} = 0.5, \epsilon = 10^{-8}, T = 80$ were used. For a given molecule, in case there are K ground truth conformations in test set, $2K$ conformations are sampled for evaluation.

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 7 of 15

In our benchmark, 8 recent or established SOTA models including GraphDG [12], ConfVAE [13], GeoMol [14], CGCF [15], ConfGF [16], GeoDiff [20], SDEGen [21], and ETKDG [18] in RDKit. The results of GraphDG, CGCF, ConfVAE, ConfGF were taken from previous study [20], while the performance of ETKDG, GeoMol, SDE-Gen and GeoDiff were evaluated by ourselves on the same test set using the provided models with various settings.

Model performance was evaluated by measuring the quality and diversity of conformations generated by different models. Here, we mainly compared four metrics built on the root mean square deviation (RMSD) proposed by Ganea et al., defined as the normalized Frobenius norm of two atomic coordinate matrices after Kabsch alignment [14]. 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 following the traditional *Recall* measure can be defined as follows:

$$COV - R\left(S_g, S_r\right) = \frac{1}{|S_r|}|C \in S_r|RMSD(C, \widehat{C} \leq \sigma, \widehat{C} \in S_g)|$$
(15)

$$MAT - R\left(S_g, S_r\right) = \frac{1}{|S_r|}\sum_{C \in S_r} \min_{\widehat{C} \in S_g}(RMSD(C, \widehat{C}))$$
(16)

where δ is a predefined threshold. The other two *Precision* based metrics, *COV-P* and *MAT-P* can be defined similarly, but with the generated and referenced set exchanged as shown in Eqs. (17, 18).

$$COV - P\left(S_g, S_r\right) = \frac{1}{|S_g|}|C \in S_g|RMSD(C, \widehat{C} \leq \sigma, \widehat{C} \in S_r)|$$
(17)

$$MAT - P\left(S_g, S_r\right) = \frac{1}{|S_g|}\sum_{C \in S_g} \min_{\widehat{C} \in S_r}(RMSD(C, \widehat{C}))$$
(18)

In practice, the S_g of each molecule is set as two times of the size of S_r. Intuitively, the *COV* score measures the percentage of structures in one set covered by another set, where covering means that the RMSD between two conformations is within a certain threshold δ. In contrast, the *MAT* score measures the average RMSD in one set with its closest neighbor in the other set. In general, a higher *COV* score or a lower *MAT* score indicates generating more realistic conformations. Moreover, the *Precision* metric measures the proportion of the true low-energy conformation among all generated conformations by the model. Here, high precision means that most generated conformation coming from the target low-energy conformation distributions within the error threshold of σ. The *Recall* measures the proportion of low-energy conformation generated by models among the true conformations in datasets. High recall means the model can generate a wide variety of low-energy conformations that cover the entire target distribution. Briefly, the *Precision* reflects more on quality while *Recall* metric reflects on diversity. Following previous works, δ is set as 0.5 Å and 1.25 Å for GEOM-QM9 and GEOM-Drugs dataset respectively.

Results and discussions

The performance of EC-Conf with different diffusion steps

The first thing to investigate is how diffusion step influences the model performance. Various models were trained by using different diffusion time steps and evaluated on a random test set. For EC-Conf, user can set the maximal diffusion steps in both forward diffusion steps and reverse iteration process, corresponding to the training and generation phase. We first evaluated the performance of EC-Conf trained with different maximal time steps of 2, 5, 10, 15, 20, 25, 50, 150 in forward ODE process, the iteration steps during generation are the same as training phase and the results are as shown in Table 1. It seems that both *COV-R* and *MAT-R* metrics reached best level when the diffusion step was set to 25. The performance on *COV-R and MAT-R* got improved when the

Table 1 Results of EC-Conf with different iteration steps on the GEMO-QM9 dataset

Training Diffusion Steps	Sampling Iteration Steps	COV-R (%)↑	MAT-R (Å)↓	COV-P (%)↑	MAT-P (Å)↓
2	2	0.335	0.565	0.468	0.532
5	5	0.459	0.522	0.857	0.356
10	10	0.759	0.361	0.868	0.340
15	15	0.763	0.342	**0.895**	**0.310**
20	20	0.792	0.354	0.873	0.352
25	25	**0.824**	**0.322**	0.863	0.337
50	50	0.805	0.339	0.828	0.354
150	150	0.503	0.514	0.413	0.566

Bold values refer to the best value

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 8 of 15

diffusion step increased from 2 to 25, while it got worse for 50 and 150. There are two possible reasons for the performance decreasing with the increase of diffusion steps. First, EC-Conf are trained by minimizing the error between adjacent steps during the forward diffusion process. During sampling, the error between adjacent iteration steps accumulates with the increase in sampling iterations, causing the final results to deviate from the ground truth structure. Second, since Ec-Conf learns the evolution relationship between diffusion time and structure, instead of only learning the relation between structure and its noise as in DDPM. Thus, an excessive number of diffusion steps increases the difficulty of learning, decreasing the overall effectiveness.

For *COV-P* and *MAT-P*, the performance of EC-Conf with 5 diffusion steps has reached 0.857 and 0.356 Å, indicating EC-Conf have learned the main features of the main low-energy conformations, but its distribution learning ability still needs improvement. The *COV-P* and *MAT-P* reached the top when the time step was set to 15, and got worse for number of iteration larger than 15. The *COV-R* increased 6% with the increase of diffusion step from 15 to 25, while *COV-P* only decreased 3%. Take

both *Recall* and *Precision* measurement into consideration, it seems that the models with diffusion step 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.

Performance metrics 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 S2. 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: Graph-DG, ConfVAE, GeoMol and iterative refinement methods: CGCF, Conf-GF, SDE-Gen, and GeoDiff. Additionally, RDKit structure fast correction (FC) option can be introduced to correct the abnormal bond lengths and angles in ML-based methods by using the bond lengths and angles of MMFF force field [37, 38]. Here, we focus on the performance of three

Table 2 The average benchmark results for EC-Conf under different sampling iterations with fixed (25) diffusion steps on GEMO-QM9 test set

Methods	Training Diffusion steps	Sampling Iteration steps	COV-R (%)↑	MAT-R (Å)↓	COV-P (%)↑	MAT-P (Å)↓
ETKDG	–	1	0.862	0.2892	0.906	0.2571
GraphDG	–	1	0.733	0.425	0.439	0.581
ConfVAE	–	1	0.778	0.415	0.38	0.622
Geomol	–	1	0.908	0.262	0.824	0.324
CGCF	1000	1000	0.78	0.422	0.365	0.662
ConfGF	5000	5000	0.885	0.267	0.464	0.522
SDEGen[a]	1500	1500	0.815	0.357	0.484	0.566
GeoDiff-A[b]	5000	5000	0.905	0.21	0.524	0.454
GeoDiff-C[c]	5000	5000	0.901	0.209	0.528	0.445
EC-Conf	25	1	0.754	0.411	0.798	0.414
		2	0.759	0.409	0.799	0.413
		5	0.830	0.348	0.832	0.378
		10	0.831	0.331	0.851	0.357
		15	0.83	0.325	0.858	0.348
		20	0.825	0.344	0.849	0.345
		25	0.824	0.322	0.863	0.337
		30	0.813	0.324	0.861	0.333
GeoDiff[b]-FC	5000	5000	0.531	0.454	0.527	0.46
SDEGen[a]-FC	1500	1500	0.559	0.484	0.473	0.525
EC-Conf-FC	25	25	0.692	0.390	0.797	0.360

[a] SDE-Gen with sampling settings: n_{euler} =250, $n_{langevin}$ =2, n_{dg}=1000

[b] GeoDiff-A trained with alignment approaches

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

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 9 of 15
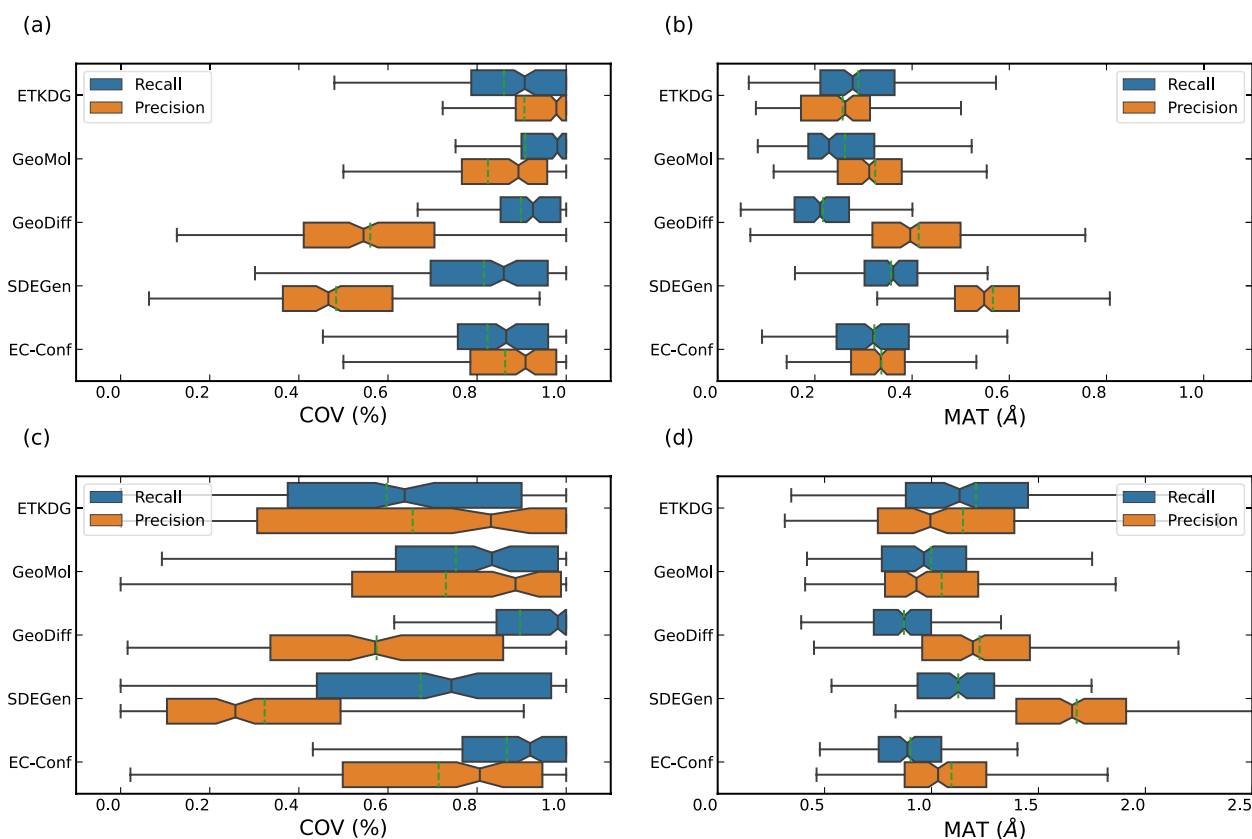
Fig. 4 The benchmarked recall and precision results of COV (**a**, **c**) and MAT (**b**, **d**) on GEOM-QM9 and GEOM-Drugs test sets, respectively

corrected diffusion models, ie. SDE-Gen-FC, Geo-Diff-FC and EC-Conf-FC.

For QM9 dataset, the recall measurement of *COV-R* and *MAT-R* are as shown in Table 2 and Fig. 4a, b, representing the diversity of generated conformations. It's clear that the diffusion-based models out-performed most one-shot models except GeoMol, suggesting their superior 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 SDE-Gen 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 2,

where the EC-Conf out-performed all the ML baselines indicating its superior quality of conformation generation under all sampling iterations. The optimal performance of EC-Conf was obtained at around 25 iterations steps on GEOM-QM9 dataset.

We also evaluate our EC-Conf models on drug-like molecules with maximum of 50 heavy atoms in GEOM-Drugs set, which is more challenging for one-shot baselines. The recall-based metrics of *COV-R* and *MAT-R* are shown in Table 3 and Fig. 4c, d. Both the GeoDiff and EC-Conf out-perform the best one-shot generative model, GeoMol. Again, even the one-shot generation of EC-Conf greatly out-performed those one-shot baselines and some of the iterative methods such as CGCF, Conf-GF, SDE-Gen. The performance of EC-Conf generation with 5 iterations already performed better than that of GeoDiff with 1000 iterations, representing 200 times better efficiency. In the concept of precision-based metrics on GEOM-Drugs set, although GeoMol out-performed the rule-based RDKit ETKDG method generally,

Table 3 The average benchmark results for EC-Conf under different sampling iterations with fixed (25) diffusion steps on GEMO-Drugs test set

Methods	Training Diffusion steps	Sampling Iteration steps	COV-R (%)↑	MAT-R (Å)↓	COV-P (%)↑	MAT-P (Å)↓
ETKDG	1	1	0.5969	1.208	0.655	1.147
GraphDG	1	1	0.083	1.972	0.021	2.434
ConfVAE	1	1	0.552	1.238	0.230	1.829
GeoMol	1	1	0.7526	0.998	0.730	1.047
CGCF	1000	1000	0.540	1.249	0.217	1.857
ConfGF	5000	5000	0.622	1.163	0.234	1.722
SDEGen[a]	1500	1500	0.560	1.237	0.251	1.720
SDEGen[b]	1500	2000	0.566	1.237	0.264	1.705
SDEGen[c]	1500	1500	0.263	1.601	0.117	2.008
SDEGen[d]	1500	6000	0.673	1.126	0.323	1.679
GeoDiff[e]	1000	1000	0.830	0.953	0.483	1.321
GeoDiff-A[f]	5000	5000	0.884	0.870	0.601	1.186
GeoDiff-C[g]	5000	5000	0.891	0.863	0.615	1.171
EC-Conf	25	1	0.797	1.023	0.705	1.118
		2	0.796	1.016	0.706	1.118
		5	0.845	0.934	0.714	1.097
		10	0.858	0.909	0.712	1.093
		15	0.859	0.905	0.716	1.084
		20	0.863	0.903	0.716	1.091
		25	0.867	0.902	0.714	1.093
		30	0.864	0.902	0.701	1.108
GeoDiff[g]-FC	5000	5000	0.797	0.953	0.606	1.189
SDEGen[d]-FC	**1500**	6000	0.776	1.017	0.567	1.237
EC-Conf-FC	25	25	0.821	0.926	0.748	1.036

[a] SDE-Gen with sampling settings: n_{euler} =250, $n_{langevin}$ =2, n_{dg}=1000

[b] SDE-Gen with sampling settings: n_{euler} =500, $n_{langevin}$ =2, n_{dg}=1000

[c] SDE-Gen with sampling settings: n_{euler} =500, $n_{langevin}$ =2, n_{dg}=500

[d] SDE-Gen with sampling settings: n_{euler} =500, $n_{langevin}$ =2, n_{dg}=5000

[e] GeoDiff trains and samples with 1000 steps

[f] GeoDiff-A trained with alignment approaches

[g] GeoDiff-C trained with chain-rule approaches

performance of EC-Conf model is quite close. Among all deep learning-based models, EC-Conf with a single iteration outperformed other models and achieved the best results with 15 sampling iterations. To intuitively demonstrate the conformation generation efficiency of EC-Conf, the conformation evolution in the sampling phase is shown in Fig. 5. It is evident that for EC-Conf, a single iteration already roughly shapes the structure, with the quality of the structure continuously improving over the next 25 steps. In contrast, for SDE-Gen and GeoDiff, the molecular structure only starts to take shape after 3500 iterations. These results further demonstrate the efficiency of EC-Conf in conformation generation.

Due to the FC option slightly adjust the ML generated structures, it decreases the model performance on both GEOM-QM9 and GEOM-Drugs test set, but it can optimize force field calculated molecular energy as discussed later. Among three corrected diffusion models, EC-Conf-FC outperforms other methods in general, only get slightly worse value on *COV-R* median as shown in Table S3.

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 11 of 15
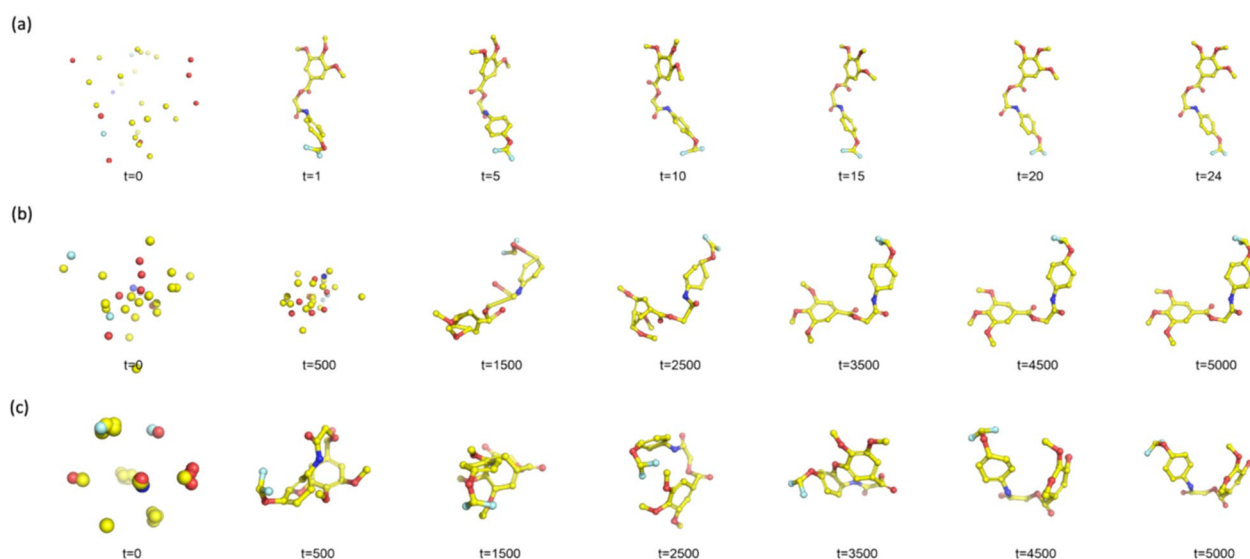
Fig. 5 The conformation evolution in **a** EC-Conf, **b** Geo-Diff diffusion process and distance geometry to coordinates phase of **c** SDE-Gen.

Some sample conformations generated by selected models are shown in Fig. 6 to provide a qualitative comparison, where EC-Conf is shown to nicely capture both local and global structures in 3D space. Additionally, SDEGen fails to generate reasonable structures for some molecules.

C. The quality of EC-Conf generated structures on GEOM-Drugs

As discussed above, machine learning methods exhibit advantages in terms of generating diverse conformations and achieving precision in reproducing ground truth conformations. Here, we evaluated the quality of EC-Conf generated conformations by examining their deviation to the optimized conformations with MMFF94 force field, and the ground truth structures with the lowest MMFF94 energy. Here, we evaluated both structural deviation and conformational energy. Comparison with structures generated by other methods, including ETKDG, SDEGen with 6000 iterations and GeoDiff model with 5000 denoising iterations. Additionally, the results from FC corrected conformations to the original optimized structure and ground truth were also compared.

The average and minimum root mean square deviation (RMSD) of the generated structures to the optimized structures on GEOM-Drugs dataset are as depicted in Fig. 7 a, b, respectively. ETKDG structures have the smallest average deviation from their optimized structures (with a median value of 0.905 Å), and the order of increasing average deviation is: ETKDG < EC-Conf < GeoDiff < EC-Conf-FC < GeoDiff-FC < SDEGen-FC < SDEGen. The structural deviation to the ground truth structures also follows the same trend as shown in Fig. 7 c, d.

Conformational energy can also serve as an indicator of structure quality, as structures with abnormal bond lengths or ring configurations can significantly increase internal energy. For methods not using FC option, GeoDiff has the smallest energy difference between generated structures and their optimized counterparts (with a median value of 0.68 kcal/mol/atom) and the order is: GeoDiff < ETKDG < EC-Conf < SDEGen as depicted in Fig. 4e. We need to note that SDEGen generates many unreasonable structures difficult to optimize, making optimized structures energy is even higher than FC corrected structures. Thus, the analysis of energy difference between SDEGen-FC and its optimized structures are meaningless. The energy difference to the ground truth follows the same trend. These findings highlight that GeoDiff model generates energy-wise most favorable conformations. However, when the FC option was used, the energy difference between generated conformation and its optimized conformation

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 12 of 15
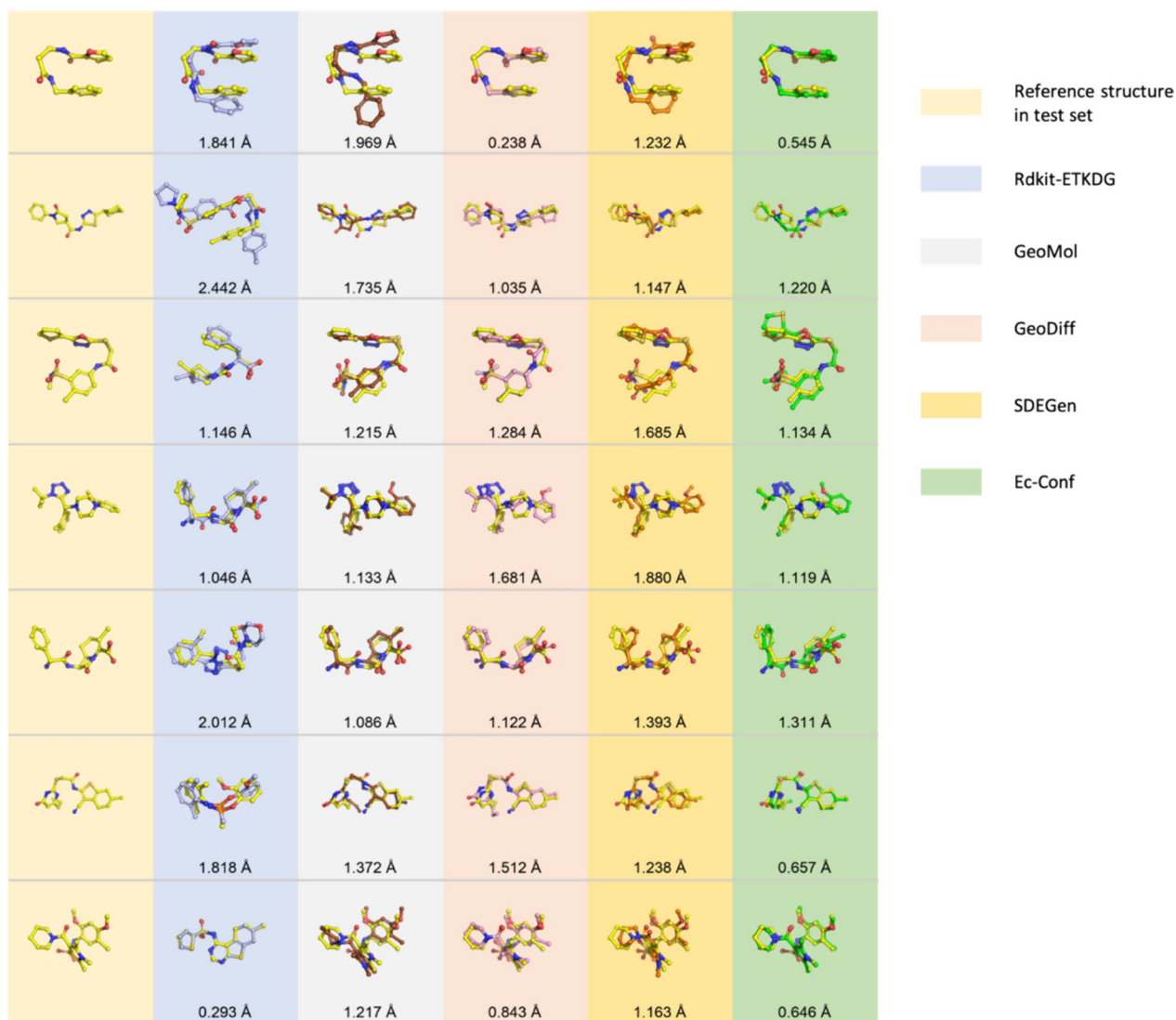
Fig. 6 Examples of generated conformations for 7 random selected molecules and their aligned RMSDs to the closest reference structures in GEOM-Drugs test set with different methods

for both EC-Conf-FC and GeoDiff-FC model decrease dramatically, which are 0.25 and 0.29 kcal/mol/atom, respectively. This suggests that user can combine EC-Conf with the structure correction functionality in RDKit to directly generate reasonable conformations without further optimization, which can save a lot of processing time considering the low diffusion step for EC-Conf. The results for GEOM-QM9 set are similar to those of GEOM-Drugs, as shown in Figure S1. All these results demonstrate that our EC-Conf model can achieve a good balance between conformation quality and sampling efficiency.
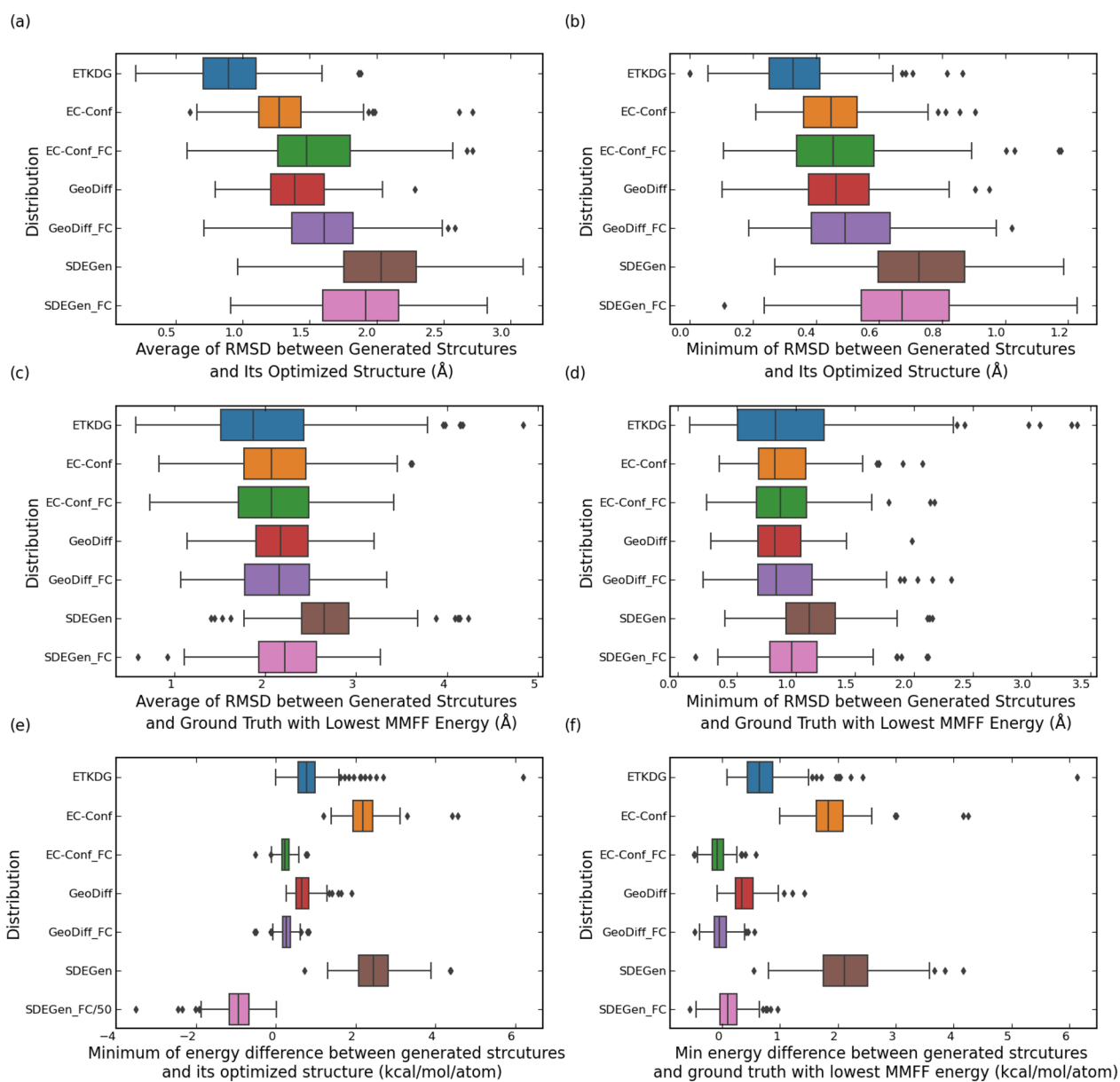
Fig. 7 Evaluation of conformation quality for various models. Average (**a**) and minimum (**b**) RMSD between generated conformations and their optimized ones. Average (**c**) and minimum (**d**) RMSD of generated structures to the ground truth structures with lowest MMFF94 energy. Minimum energy difference of generated structure to its optimized structures (**e**) and the ground truth structure with lowest MMFF94 energy (**f**). *SDEGen-FC/50 means scaling the result by 50 times

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 14 of 15

Conclusions

Here, an equivariant consistency generative 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.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13321-024-00893-2.

> Supplementary file1 (DOCX 530 KB) 1. The algorithm of equivariant consistency training and sampling is as shown in Algorithm S1 and S2. 2. The median of benchmark results and 7 conformation examples on GEOM-QM9 are as shown in Table S1 and Figure S1. The median of benchmark result on GEOM-Drugs is as shown in Table S2. 3. Quality of generations on GEOM-QM9 Datasets is as shown in Figure S2.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Author details

^{1}School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou 510006, China. ^{2}Guangzhou National Laboratory, Guangzhou 510005, China. ^{3} Guangzhou Medical University, Guangzhou, 511495, China.

References

1. Perola E, Charifson PS (2004) Conformational analysis of drug-like molecules bound to proteins: an extensive study of ligand reorganization upon binding. J Med Chem 47(10):2499–2510
2. Lyne PD (2002) Structure-based virtual screening: an overview. Drug Discov Today 7(20):1047–1055
3. Jain AN (2004) Ligand-based structural hypotheses for virtual screening. J Med Chem 47(4):947–961
4. Bhunia SS, Saxena M, Saxena AK. Ligand-and structure-based virtual screening in drug discovery. In: Biophysical and Computational Tools in Drug Discovery. Springer; 2021. p. 281–339.
5. Broccatelli F, Brown N (2014) Best of both worlds: on the complementarity of ligand-based and structure-based virtual screening. J Chem Inf Model 54(6):1634–1641
6. Cruciani G, Carosati E, Clementi S. Three-dimensional quantitative structure-property relationships; 2003.
7. Schwab CH (2010) Conformations and 3D pharmacophore searching. Drug Discov Today Technol 7(4):e245–e253
8. Hendrickson MA, Nicklaus MC, Milne GW, Zaharevitz D (1993) CONCORD and CAMBRIDGE: comparison of computer generated chemical structures with x-ray crystallographic data. J Chem Inf Comput 33(1):155–163
9. Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. Tetrahedron Comput Methodol 3(6):537–547
10. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. J Chem Inf Model 50(4):572–584
11. Mansimov E, Mahmood O, Kang S, Cho K (2019) Molecular geometry prediction using a deep generative graph neural network. Sci Rep 9(1):20381
12. Simm GN, Hernández-Lobato JM. A generative model for molecular distance geometry. In: ICML 2020. PMLR 2020.
13. Xu M, Wang W, Luo S, Shi C, Bengio Y, Gomez-Bombarelli R, Tang J. An end-to-end framework for molecular conformation generation via bilevel programming. In: ICML 2021: 2021. PMLR: 11537–11547.
14. Ganea O, Pattanaik L, Coley C, Barzilay R, Jensen K, Green W, Jaakkola T. Geomol: Torsional geometric generation of molecular 3d conformer ensembles. In: NIPS 2021. 2021: 13757–13769.
15. Xu M, Luo S, Bengio Y, Peng J, Tang J. Learning neural generative dynamics for molecular conformation generation. In: ICLR 2021. 2021.
16. Shi C, Luo S, Xu M, Tang J. Learning gradient fields for molecular conformation generation. In: ICML 2021: 2021. PMLR: 9558–9568.
17. Schärfer C, Schulz-Gasch T, Ehrlich H-C, Guba W, Rarey M, Stahl M (2013) Torsion angle preferences in druglike chemical space: a comprehensive guide. J Med Chem 56(5):2016–2028
18. Wang S, Witek J, Landrum GA, Riniker S (2020) Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. J Chem Inf Model 60(4):2044–2058
19. Zhu J, Xia Y, Liu C, Wu L, Xie S, Wang Y, Wang T, Qin T, Zhou W, Li H: Direct molecular conformation generation. In: Transactions on Machine Learning Research. 2022.
20. Xu M, Yu L, Song Y, Shi C, Ermon S, Tang J. Geodiff: A geometric diffusion model for molecular conformation generation. In: ICLR 2022. 2022.
21. Zhang H, Li S, Zhang J, Wang Z, Wang J, Jiang D, Bian Z, Zhang Y, Deng Y, Song J et al (2023) SDEGen: learning to evolve molecular conformations from thermodynamic noise for conformation generation. Chem Sci 14(6):1557–1568
22. Jing B, Corso G, Chang J, Barzilay R, Jaakkola T. Torsional diffusion for molecular conformer generation. In: NIPS 2022. 2022: 24240–24253.
23. Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. NIPS 2020:6840–6851
24. Song Y, Ermon S. Generative modeling by estimating gradients of the data distribution. In: NIPS 2019. 2019.
25. Song Y, Ermon S. Improved techniques for training score-based generative models. In: NIPS 2020. 2020: 12438–12448.
26. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. In: ICLR 2021. 2021.

Fan *et al. Journal of Cheminformatics*   (2024) 16:107

Page 15 of 15

27. Song Y, Dhariwal P, Chen M, Sutskever I. Consistency models. In: ICML 2023. 2023.
28. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA (2018) Generative adversarial networks: An overview. IEEE Signal Process Mag 35(1):53–65
29. Kingma DP, Welling M. Auto-Encoding Variational Bayes. 2013: arXiv:1312.6114.
30. Jimenez Rezende D, Mohamed S. Variational Inference with Normalizing Flows. In: *ICML 2015: May 01, 2015*. 2015: arXiv:1505.05770.
31. Yang L, Zhang Z, Song Y, Hong S, Xu R, Zhao Y, Shao Y, Zhang W, Cui B, Yang M-H. Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:220900796 2022.
32. Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. In: ICLR 2021. 2020.
33. Karras T, Aittala M, Aila T, Laine S. Elucidating the design space of diffusion-based generative models. In: NIPS 2022. 2022: 26565–26577.
34. Liao Y-L, Smidt T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In: ICLR 2023. 2022.
35. Ramakrishnan R, Dral PO, Rupp M, Von Lilienfeld OA (2014) Quantum chemistry structures and properties of 134 kilo molecules. Sci Data 1(1):1–7
36. Axelrod S, Gomez-Bombarelli R (2022) GEOM, energy-annotated molecular conformations for property prediction and molecular generation. Sci Data 9(1):185
37. Zhang X, Zhang O, Shen C, Qu W, Chen S, Cao H, Kang Y, Wang Z, Wang E, Zhang J et al (2023) Efficient and accurate large library ligand docking with KarmaDock. Nat Comput Sci 3(9):789–804
38. Stärk H, Ganea O, Pattanaik L, Barzilay R, Jaakkola T: Equibind. Geometric deep learning for drug binding structure prediction. In: ICML 2022: 2022. PMLR: 20503–20521.

Publisher's Note