中國人民大學
RENMIN UNIVERSITY OF CHINA

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Align Your Structures: Generating Trajectories with Structure Pretraining for Molecular Dynamics

Aniketh Iyengar, Jiaqi Han, Pengwei Sun, Mingjian Jiang,
Jianwen Xie, Stefano Ermon

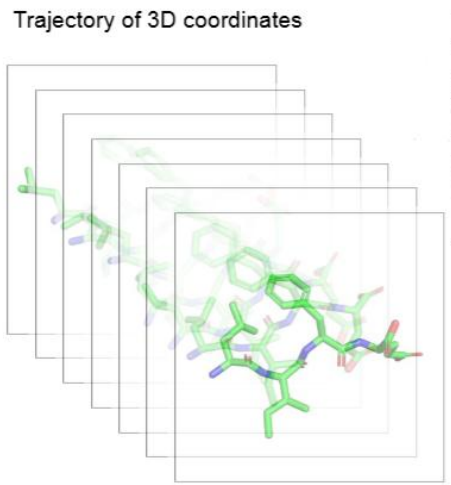
ICML 2025 Workshop GenBio

Fanmeng Wang

2025-10-20

Introduction

- **Molecular Dynamics (MD)** is a typical computational simulation method used to model the physical movements of atoms and molecules over time.
- Despite its utility, MD is computationally intensive, often requiring long simulation times and a large number of small integration steps to ensure numerical precision.
- In this context, using deep generative models to generate molecular dynamics (MD) trajectories has attracted increasing attention.



Note: The MD trajectories only contain post-minimization dynamics, representing the system's real thermal motion on the equilibrated potential surface.

Introduction

- However, existing MD generative models are typically optimized on a single or a group of limited number of molecular systems, making it a fundamental challenge for them to generalize across different molecules.

  This is primarily due to two reasons:

  - **Data scarcity**: Building large-scale MD dataset over diverse molecular systems is very prohibitive, leading to insufficiency in the amount of data for the models to well capture the underlying MD distribution.
  - **Modeling complexity**: MD data is of high-dimensionality by extending the molecular structure space with an additional temporal dimension, which further contributes to modeling difficulty.

Introduction

■ EGINTERPOLATOR

➢ To overcome these challenges, we propose a novel framework that leverages the two-stage pipeline for MD trajectory generation.

- **Conformer pretraining**: Generate individual frames (i.e., static molecular structures / conformers) along each MD trajectory.
- **Temporal alignment**: Capture temporal consistency along each MD trajectory.

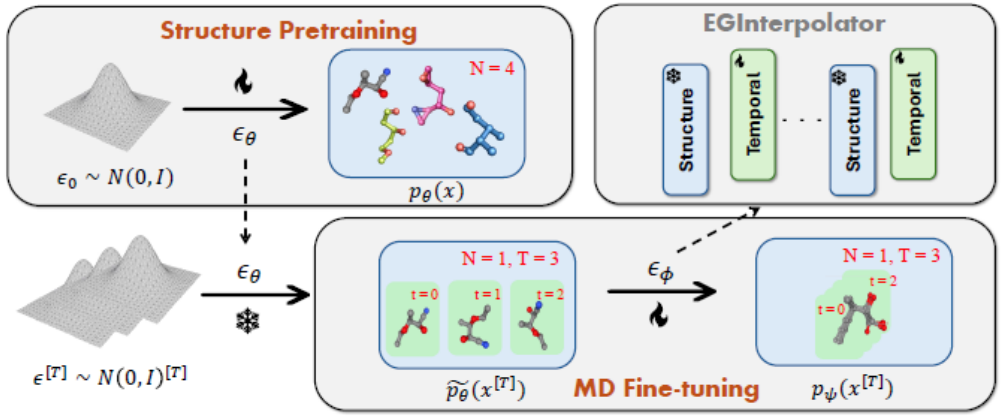


Figure 1. The overall framework of EGINTERPOLATOR.

Method

Notations

- **Each molecular dynamics trajectory** is represented as a collection of static structures (i.e., conformers) that evolve through time.
- **Each frame** at timestep t is represented as a geometric graph $(\mathbf{h}, \mathbf{x}^{(\mathbf{t})}, \mathcal{E})$:
  - $\mathbf{h_i} \in R^H$ refers to the node feature of atom
  - $\mathbf{x}^{(\mathbf{t})}$ refers to the 3D coordinate of atom i at timestep t
  - $\mathcal{E}$ refers to the set of edges induced by the chemical bonds between atoms
- In this context, the MD trajectory with length T is represented as:

$$\mathbf{x}^{[\mathbf{T}]} := \mathbf{x}^{(\mathbf{0}:\mathbf{T}-\mathbf{1})} \in R^{T \times N \times 3}$$

Method

■ Stage 1: Conformer pretraining

- We firstly pretrain a conformer diffusion model on large-scale conformer datasets, learning to generate conformers—i.e., the static molecular structures corresponding to individual frames along an MD trajectory.
- In particular, we employ **Equivariant Graph Convolution Layer (EGCL)** as the **E**quivariant **S**tructure layer, whoe update is denoted as follows:

$$\mathbf{x}', \mathbf{h}' = f_{\mathrm{ES}}(\mathbf{x}, \mathbf{h}, \mathcal{E}),$$

where the **denoiser** $\boldsymbol{\epsilon}_\theta$ consists of L layers of f_{ES} stacked sequentially and is then optimized using the following loss:

$$\mathcal{L}_{\mathrm{struct}} = \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}_{\mathrm{struct}}, \tau, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\mathbf{x}_\tau, \tau) \right\|_2^2,$$

Method

■ Stage 2: Temporal alignment

➢ We propose a temporal interpolator module that entangles the pretrained structure denoiser ϵ_θ with the additional temporal network $\epsilon_\emptyset$ through a linear interpolation, learning to capture temporal consistency along each MD trajectory.

$$\begin{aligned}\epsilon'_\psi(\mathbf{x}_\tau^{[T]}, \tau) =& \alpha\tilde{\mathbf{x}}_\tau^{[T]} + (1-\alpha)\epsilon_\phi(\tilde{\mathbf{x}}_\tau^{[T]}, \tau), \\ \text{s.t. } \tilde{\mathbf{x}}_\tau^{[T]} =& [\epsilon_\theta(\mathbf{x}_\tau^{(t)}, \tau)]_{t=0}^{T-1},\end{aligned}$$

➢ In particular, we employ **Equivariant Temporal Attention Layer** as **E**quivariant **T**emporal layer , whoe update is denoted as follows:

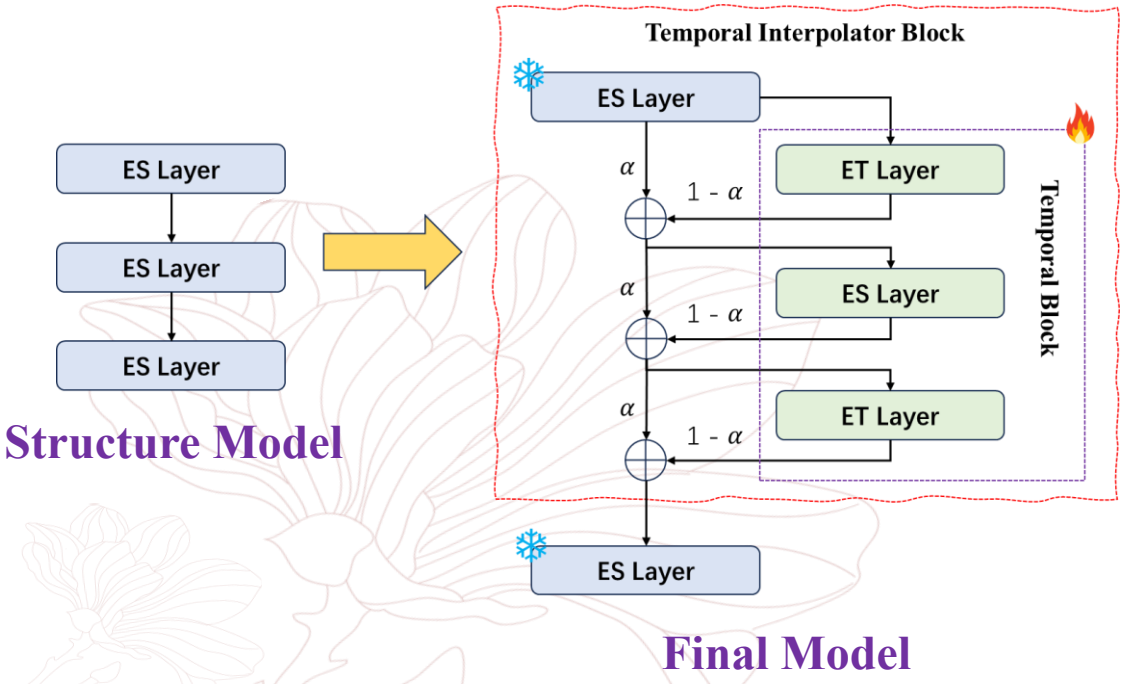
$$\mathbf{x}'^{[T]}, \mathbf{h}'^{[T]} = f_{\mathrm{ET}}(\mathbf{x}^{[T]}, \mathbf{h}^{[T]}, \mathcal{E}),$$

where **each temporal block** as a stack of three layers, with one ET layer on the top, one on the bottom, and an ES layer in the middle.

Method

■ Stage 2: Temporal alignment

➢ **Model architecture**: For each ES layer in the pretrained structure model, we initialize one temporal block, which, by putting together, constitutes one temporal interpolator block, leading to L temporal interpolator blocks in total.



➢ During training, we keep the pretrained ES layers freezed, and the model is optimized using the following loss:

$$\mathcal{L}_{\text{traj}} = \mathbb{E}_{\mathbf{x}_0^{[T]} \sim \mathcal{D}_{\text{traj}}, \tau, \boldsymbol{\epsilon}^{[T]}} \| \boldsymbol{\epsilon}^{[T]} - \boldsymbol{\epsilon}'_\psi(\mathbf{x}_\tau^{[T]}, \tau) \|_2^2,$$

➢ Here, this final model is not only a performant MD generative model, but also yields exactly no performance degradation on the conformer generation task (just setting α as 1).

Experiment

Datasets

- **Conformer datasets**: We employ GEOM-QM9 and GEOM-Drugs for conformer pretraining, while the structure model is pretrained separately on each dataset.
- **MD datasets:** We subsample QM9 and Drugs datasets, then perform five, all-atom, explicit-solvent simulations of 5 ns per molecule.

Metrics

- **Conformer generation**: To compare the performance of our pretrained structure model on the conformer generation task, we use the Coverage and Matching metrics, reporting both the Recall and Precision.
- **MD generation**: We assess the similarity between generated and reference trajectories using average Jensen–Shannon divergence (JSD) across key collective variable distributions, including bond lengths, bond angles, torsion angles and the top components from time-lagged independent component analysis (TICA).

Experiment

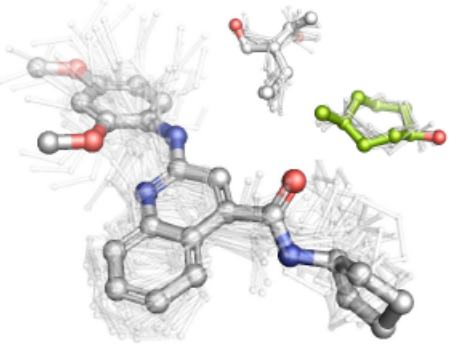
■ Conformer generation

➢ Our pretrained structure model performs competitively with SOTA methods.

A. Coverage and Matching Results on QM9 and GEOM-Drugs

	Method	COV-R (%) ↑		MAT-R (Å) ↓		COV-P (%) ↑		MAT-P (Å) ↓	
		Mean	Med.	Mean	Med.	Mean	Med.	Mean	Med.
QM9	CONFGF	88.49	94.31	0.2673	0.2685	46.43	43.41	0.5224	0.5124
	GEODIFF-A	90.54	94.61	0.2104	0.2021	52.35	50.10	0.4539	0.4399
	BASICES	80.38	82.12	0.2819	0.2941	58.83	55.13	0.4298	0.4230
Drugs	CONFGF	62.15	70.93	1.1629	1.1596	23.42	15.52	1.7219	1.6863
	GEODIFF-A	88.36	96.09	0.8704	0.8628	60.14	61.25	1.1864	1.1391
	BASICES	93.15	100.00	0.7932	0.7812	69.68	76.35	1.0837	1.0381

B. Generated Conformers



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

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

Experiment

■ MD generation

➢ Our model achieves SOTA performance compared with existing methods.

Table 1. Performance Comparison on QM9 Unconditional Generation and Drugs Forward Simulation

	Method	JSD (Mean — Median) (↓)									
		Bond Angle		Bond Length		Torsion		TICA_0		TICA_0,1	
QM9	MD ORACLE	0.042	0.028	0.032	0.031	0.192	0.134	0.318	0.291	0.413	0.394
	AR + EGNN	0.702	0.677	0.770	0.780	0.702	0.761	0.770	0.788	0.820	0.824
	AR + ET	0.705	0.746	0.680	0.721	0.553	0.586	0.568	0.562	0.783	0.786
	GEOTDM	0.691	0.690	0.676	0.670	0.489	0.527	0.449	0.453	0.691	0.694
	EGINTERPOLATOR	**0.305**	**0.292**	**0.210**	**0.188**	**0.363**	**0.380**	**0.417**	**0.406**	**0.636**	**0.642**
Drugs	MD ORACLE	0.036	0.023	0.030	0.028	0.215	0.131	0.484	0.494	0.610	0.630
	AR + EGNN	0.663	0.655	0.748	0.784	0.723	0.741	0.716	0.731	0.806	0.821
	AR + ET	0.765	0.766	0.733	0.745	0.526	0.533	0.565	0.558	0.791	0.795
	GEOTDM	0.640	0.645	0.643	0.645	0.498	0.503	0.531	0.550	0.712	0.720
	EGINTERPOLATOR	**0.173**	**0.153**	**0.1419**	**0.112**	**0.377**	**0.388**	**0.454**	**0.441**	**0.650**	**0.644**

Unconditional generation: Generate trajectories with no reliance on a reference frame

Forward simulation: Generate trajectories conditioned on the first frame

Experiment

■ Ablations

➢ Structural Pretraining

- The results highlight that structural pretraining enriches limited dynamic data and facilitates learning of accurate spatiotemporal distributions.

Table 2. Ablation study on QM9 and Drugs.

	Method	JSD (Mean — Median) (↓)							
		Bond Angle		Bond Length		Torsion		Decorrelation	
		Mean	Median	Mean	Median	Mean	Median	Mean	Median
QM9	OURS-N	0.538	0.538	0.583	0.580	0.441	0.494	0.619	0.718
	OURS	**0.305**	**0.292**	**0.210**	**0.188**	**0.363**	**0.380**	0.607	0.727
Drugs	OURS-N	0.332	0.332	0.386	0.383	0.455	0.466	0.720	0.833
	OURS	**0.173**	**0.153**	**0.142**	**0.112**	**0.377**	**0.388**	**0.670**	**0.794**

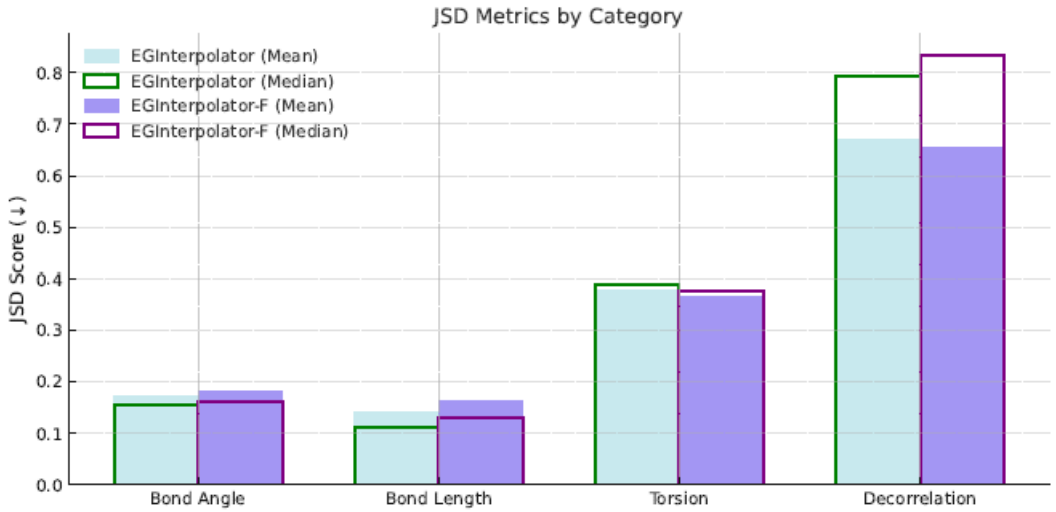
EGINTERPOLATOR-N: Trained directly on trajectories without any conformer pretraining.

Experiment

■ Ablations

➢ Frozen ES Layers

- The performance remains largely unchanged across metrics, indicating that the pretrained structure model generalizes well without task-specific tuning, while temporal layers effectively capture the necessary dynamic information.



EGINTERPOLATOR-F: Trained fully end-to-end on trajectories without freezed

Conclusion

- We propose a two-stage framework for modeling MD distributions by pretraining a structure model on conformer dataset and then finetuning on trajectory dataset.
- By **mixing the output** from the pretrained structure model and the temporal model to capture the temporal dependency, our method demonstrates strong performance in terms of producing realistic MD trajectories on diverse benchmarks and tasks.
- **Limitation:** It's still difficult to yield MD simulation that is at the same level of accuracy as the MD (i.e., those generated through MD).

Thank You for listening!

Fanmeng Wang

2025-10-20