LagNet: Deep Lagrangian Mechanics for Plug-and-Play Molecular Representation Learning

Chunyan Li[1,2], Junfeng Yao[1], Jinsong Su[1], Zhaoyang Liu[3], Xiangxiang Zeng[4*], Chenxi Huang[1†]

[1] School of Informatics, Xiamen University, Xiamen, China
[2] School of Informatics, Yunnan Normal University, Kunming, China
[3] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China
[4] College of Computer Science and Electronic Engineering, Hunan University, Changsha, China
lchy0316@gmail.com, xzeng@hnu.edu.cn, supermonkeyxi@xmu.edu.cn

Abstract

Molecular representation learning is a fundamental problem in the field of drug discovery and molecular science. Whereas incorporating molecular 3D information in the representations of molecule seems beneficial, which is related to computational chemistry with the basic task of predicting stable 3D structures (conformations) of molecules. Existing machine learning methods either rely on 1D and 2D molecular properties or simulate molecular force field to use additional 3D structure information via Hamiltonian network. The former has the disadvantage of ignoring important 3D structure features, while the latter has the disadvantage that existing Hamiltonian neural network must satisfy the "canonial" constraint, which is difficult to be obeyed in many cases. In this paper, we propose a novel plug-and-play architecture LagNet by simulating molecular force field only with parameterized position coordinates, which implements Lagrangian mechanics to learn molecular representation by preserving 3D conformation without obeying any additional restrictions. LagNet is designed to generate known conformations and generalize for unknown ones from molecular SMILES. Implicit positions in LagNet are learned iteratively using discrete-time Lagrangian equations. Experimental results show that LagNet can well learn 3D molecular structure features, and outperforms previous state-of-the-art baselines related molecular representation by a significant margin.

Introduction

Well-designed representations of molecules are the basis of drug design and molecular property prediction. Most of the previous deep learning-based molecular representation studies focused on 1D sequence manner using Simplified Molecular Input Line Entry System (SMILES) (Li et al. 2022a), molecular fingerprint (Rogers and Hahn 2010) and 2D graph (Kearnes et al. 2016) or a mixture architecture of 1D and 2D property features (Paul et al. 2018). These methods generally used off-the-shelf neural architectures such as graph convolutional network (GCN) (Kipf and Welling 2016), Message-Passing Neural Networks (MPNNs) (Gilmer et al. 2017), recurrent neural network (RNN) (Li et al. 2022a) and Transformer (Rong et al. 2020a) to learn molecular representation. However, these methods ignore important 3D structural features of molecules. Recently, there has been a trend towards incorporating 3D structural features of molecules into molecular representation for various downstream tasks (Li et al. 2021b,a). Whereas it is difficult to obtain labeled 3D information of molecules. The molecular conformation (the 3D coordinates of atoms in a molecule) contains important geometric information and plays an important role in the prediction of biochemical function and activity. The 3D structural features of a molecule largely determine the properties of drugs and the binding features of drug targets. It has been shown that using the 3D coordinates of atoms in 3D space can improve the accuracy of molecular property predictions (Schütt et al. 2017; Cho and Choi 2018; Li et al. 2021c; Yang et al. 2021). 3D structure is one of the most critical factors in determining molecular properties and understanding how they function in the physical world (Wu et al. 2021).

To push the boundaries of molecular characterization into the 3D realm, we propose LagNet from the perspective of molecular dynamics simulations. From physics, molecules are considered as particle systems, the motion of all particles follows the classical Newton's laws of motion, and the interaction between atoms satisfies the superposition principle. The predominant conformations of the molecules reflect the equilibrium of these particle systems and are therefore of great interest. Inspired by HamNet (Li et al. 2021c) and PhysChem (Yang et al. 2021), in this paper, we proposed LagNet, a plug-and-play molecular representation method based on Lagrangian mechanism. In summary, the key contributions of our work are as follows:

- To the best of our knowledge, we are the first to propose a Lagrangian-based mechanism for modeling small molecule conformations using deep neural network with only parameterized positional coordinates to simulate molecular dynamics processes. LagNet learns molecular representations by maintaining 3D conformations without adhering to additional "canonial" constraints that are difficult to obey in deep learning networks.
- Taking SMILES strings of molecules as input, LagNet is designed to generate known molecular conformations and can be generalized to unknown molecular conforma-

*Corresponding author: Xiangxiang Zeng
†Corresponding author: Chenxi Huang

tions through the discrete-time Lagrange equations. The potential energy and atomic momentum of molecules are updated by iteratively learning the implicit position, and the final positions predicted by neural network are supervised with the real molecular conformation information.

- Extensive experiments are conducted on different datasets to demonstrate the effectiveness of LagNet. Our molecular experiments for reconstructing conformation show that the proposed LagNet can predict molecular conformations more accurately than neural physical engines (HamNet, PhysChem), traditional geometric methods (RDKit), and message-passing neural architectures (MPNN). Meanwhile, LagNet learns conformations of molecules from scratch, therefore, labeled molecular conformations are not required. Moreover, LagNet can accurately predict the 3D conformation structure of any molecule.

Related Work

Molecular Representation Learning

Early molecular representation learning encoded SMILES strings, molecular fingerprint and graph notations using off-the-shelf neural architectures. The graph-based neural networks, such as GCN (Kipf and Welling 2016), can directly process graph topology and vertex attribute information, extract features from molecular graph structure (Hao et al. 2020). MPNNs (Gilmer et al. 2017) introduced a general framework for graph-based model, that includes message passing process and readout phases. The former aims to extract molecular graph features, and the latter is to obtain molecular graph-level representation for downstream tasks. Whereas incorporating molecular 3D information in the representations of molecule seems beneficial (Schütt et al. 2017; Klicpera, Groβ, and Günnemann 2020; Cho and Choi 2018). The models that used geometric features of 3D molecular conformations achieved better performance than those that used only 1D or 2D features notations. However, above molecular 3D models required labeled conformations, which is difficult to obtain. The common practice of most models is to generate molecular 3D structure by using the molecular force field in the RDKit package (Landrum 2010), which has large noise and leads to inaccurate 3D coordinates. Therefore, it limited the applicability of 3D representation models.

Molecular Dynamics Simulation

Recent studies demonstrated that neural networks successfully learn physical dynamics to simulate potential energy, kinetic energy and forces in a system of particles, which facilitates fast molecular simulations (Zhang et al. 2018; Lu et al. 2021), accelerating the progress of conventional molecular dynamics. More recently, researchers began to simulate molecular dynamics to predict 3D conformation using neural networks, which includes two categories: one is to predict conformation distribution by flow-based and energy-based models (Xu et al. 2021b), or to generate conformations by gradient field and scoring function methods (Xu et al. 2021a; Luo et al. 2021), or to infer the 3D coordinates of the atoms by diffusion model and coarse-graining probabilistic model (Xu et al. 2022; Wand et al. 2022). The second category is to generate 3D conformational information by a neural physical engine (simulating molecular force field) using Hamiltonian neural network (Li et al. 2021c; Yang et al. 2021). HamNet (Li et al. 2021c) reconstructed molecular conformations using Hamiltonian engine with parameterized potential, kinetic and dissipation functions. PhysChem (Yang et al. 2021) fused physical and chemical information of molecules to learn molecular representations, which directly parameterize the forces between each pair of atoms. The disadvantage of the first category is the extra errors induced by the distance prediction. In addition, when generating 3D coordinates, the relationship of three sides and interior angle summation theorem in a triangle cannot be satisfied. The disadvantage of the second category is that the existing Hamiltonian neural network must satisfy the "canonial" constraint, which is difficult to be obeyed in many cases. Specifically, considering the Hamiltonian $\mathcal{H}$ in a system, the Hamiltonian formalism requires to obey a strict set of rules given by the Poisson bracket relations via the input coordinates (p, q): $\{\mathcal{H}, \mathcal{H}\} = 0$, where $\{F, G\}_{pq} = \sum_t (\frac{\partial F}{\partial p_t}\frac{\partial G}{\partial q_t} - \frac{\partial F}{\partial q_t}\frac{\partial G}{\partial p_t})$. In addition, some discrete gradient methods were implemented to model the dynamics of physical systems (Matsubara, Ishikawa, and Yaguchi 2020; Aoshima, Matsubara, and Yaguchi 2021), such as pendulum system and mass-spring systetm. However, the approaches are not directly applicable to molecular representation learning.

Preliminaries

Hamiltonian equations. The Hamiltonian equations describe the relationship between particle momentum, particle space position and Hamiltonian in a multi-particle system. In a system with n particles, the position and momentum of n particles can be denoted as $(x_1, x_2, ..., x_n)$ and $(p_1, p_2, ..., p_n)$, respectively. x_i and p_i are regarded as independent variables. Let $\mathcal{H}$ is the Hamiltonian of the system, so the Hamiltonian equations can be expressed as follows:

$$\frac{dp_i}{dt} = -\frac{\partial \mathcal{H}}{\partial x_i}, \frac{dx_i}{dt} = \frac{\partial \mathcal{H}}{\partial p_i} \tag{1}$$

The total energy of the system is expressed by Hamiltonian $\mathcal{H}$ as follows, which is the sum of the kinetic energy T of each particle and the potential energy V indicated the interaction of the system:

$$\mathcal{H} = \sum_{i=1}^{n} T_i(p_i) + V(x_1, x_2, ..., x_n) \tag{2}$$

Lagrangian mechanics and discrete gradient. Let $\mathcal{L}$ be the Lagrangian, and $\mathcal{L} = T - V$, where T denotes kinetic energy and V denotes potential energy. If M is the mass-matrix, $\ddot{x} = \frac{d^2x}{dt^2}$ indicates the second derivative of x with respect to time t, then $M\ddot{x}$ indicates the force f that atoms experience, and $f = -\nabla_x V(x)$, so $M\ddot{x} = -\nabla_x V(x)$. The Euler–Lagrangian equation can be denoted as $\frac{d}{dt}\nabla_{\dot{x}}\mathcal{L} = \nabla_x \mathcal{L}$ (Cranmer et al. 2020; Lutter, Ritter, and Peters 2019).

Some studies (Celledoni et al. 2012; Matsubara, Ishikawa, and Yaguchi 2020; Aoshima, Matsubara, and Yaguchi 2021) employed discrete gradient methods to conserve physical system energy strictly.

GCNs and MPNNs. Given a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ with initial node features f_{x_v} for $v \in \mathcal{V}$ and edge features $f_{e_{uv}}$ for $(u, v) \in \mathcal{E}$. $|v| = n$, the adjacency matrix $A \in \mathcal{R}^{n \times n}$, and $\hat{A} = A + I_n$, I_n denotes an identity matrix. The degree matrix $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$. GCN (Kipf and Welling 2016) extracts the hidden state features of nodes as

$$H^{(l+1)} = f(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{3}$$

Here, $W^{(l)}$ is a trainable weight matrix. $f(\cdot)$ denotes an activation function. $H^{(l)}$ is the matrix of activations in the l^{th} layer and $H^{(0)}$ denotes the original feature matrix. MPNNs (Gilmer et al. 2017) can obtain nodes representations h_v for every hidden layer via learning the graph structures and passing message. Calculate the m-th iteration of message passing follows

$$h_v^{(m)} = Y_U^{(m)}(h_v^{(m-1)}, Y_A^{(m)}(\{(h_v^{(m-1)}, h_u^{(m-1)}, f_{e_{uv}})\})) \tag{4}$$

where $u \in \mathcal{N}(v)$, $\mathcal{N}(v)$ is the neighbor set of node v, $h_v^{(m)}$ is the feature vector of node v at the m-th layer, $f_{e_{uv}}$ is the embedding of the edge between node u and node v, $h_v^{(0)} = f_{x_v}$. $Y_U(\cdot)$ and $Y_A(\cdot)$ are parameterized by neural networks. $Y_A(\cdot)$ denotes an aggregate function and $Y_U(\cdot)$ denotes an update function.

Lagrangian Molecular Representation: LagNet

Figure 1 shows the overview of LagNet. LagNet learns the system dynamics of molecular conformations and is designed to be a plug-and-play molecular representation module, consisting of four parts: featurization, position initialization, Lagrangian molecular dynamics and graph attention layer for downstream tasks. Inspired by mass-spring system (Aoshima, Matsubara, and Yaguchi 2021), the proposed LagNet is a discrete-time Lagrangian mechanics network for molecular representation learning, which can simulate atomic interactions in a physical system of molecule.

Discrete Lagrangian equations. From Eq. (1),

$$\frac{dp}{dt} = -\nabla_x \mathcal{H}, \frac{dx}{dt} = \nabla_p \mathcal{H} \tag{5}$$

Here, p denotes atomic momentum in a molecule, and x denotes atomic position coordinate. From Eq. (2), Hamiltonian $\mathcal{H}$ is a linear combination of kinetic energy T and potential energy V. T and V are parameterized by momentum p and position x, respectively. Specifically, T can be expressed as:

$$T = \frac{1}{2}p^{\intercal}M^{-1}p \tag{6}$$

Therefore , $\frac{dp}{dt}$ and $\frac{dx}{dt}$ can be calculated (Zhong, Dey, and Chakraborty 2020):

$$\frac{dp}{dt} = -\nabla_x V(x), \frac{dx}{dt} = M^{-1}p \tag{7}$$

Let $\overline{\nabla}$ denotes discrete gradient. A vector $\overline{\nabla}J$ is a discrete gradient of a function J when $\overline{\nabla}J$ satisfies the following

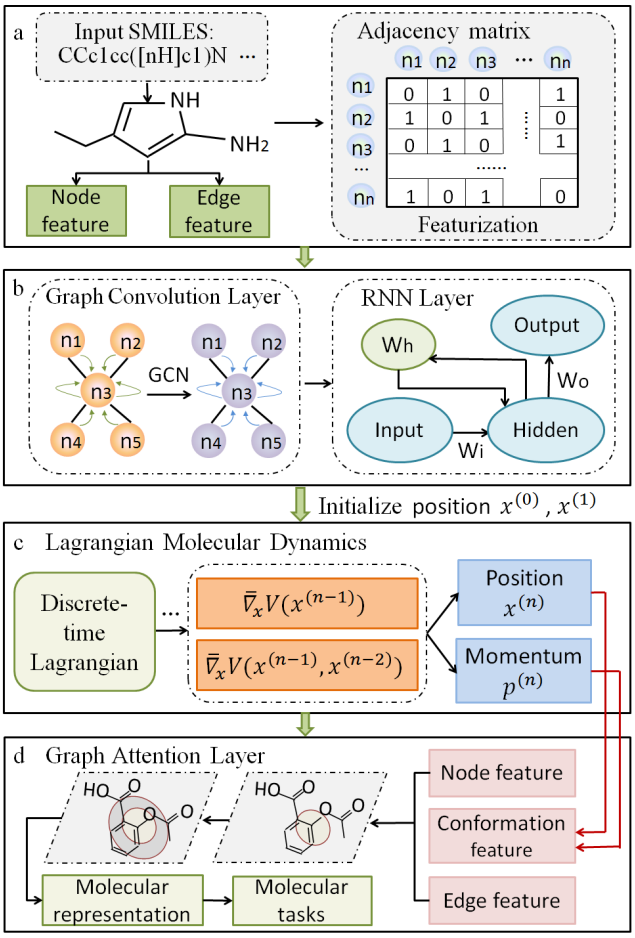


Figure 1: Overview of LagNet, a deep Lagrangian mechanics for plug-and-play molecular representation learning. a. The feature engineering to obtain node feature, edge feature and adjacency matrix by converting molecular SMILES strings into molecular graph. b. The position initialization by using graph convolution network and recurrent neural network. c. The process of Lagrangian molecular dynamics for predicting atomic position and momentum. d. The graph attention layer for molecular tasks with the input of node feature, edge feature and conformation feature using atomic position and momentum.

condition:

$$J(m) - J(n) = \overline{\nabla}J(m,n) \cdot (m-n), \overline{\nabla}J(m,m) = \nabla J(m) \tag{8}$$

where $\cdot$ indicates an operation of inner product. Following discrete gradient methods (Aoshima, Matsubara, and Yaguchi 2021; Matsubara, Ishikawa, and Yaguchi 2020), a discrete gradient $\overline{\nabla}\mathcal{H}$ for Eq. (5) can be expressed as follows:

$$\frac{p^{(n+1)} - p^{(n)}}{\Delta t} = -\overline{\nabla}_x \mathcal{H}(z^{(n+1)}, z^{(n)}) \tag{9}$$

$$\frac{x^{(n+1)} - x^{(n)}}{\Delta t} = \overline{\nabla}_p \mathcal{H}(z^{(n+1)}, z^{(n)}) \tag{10}$$

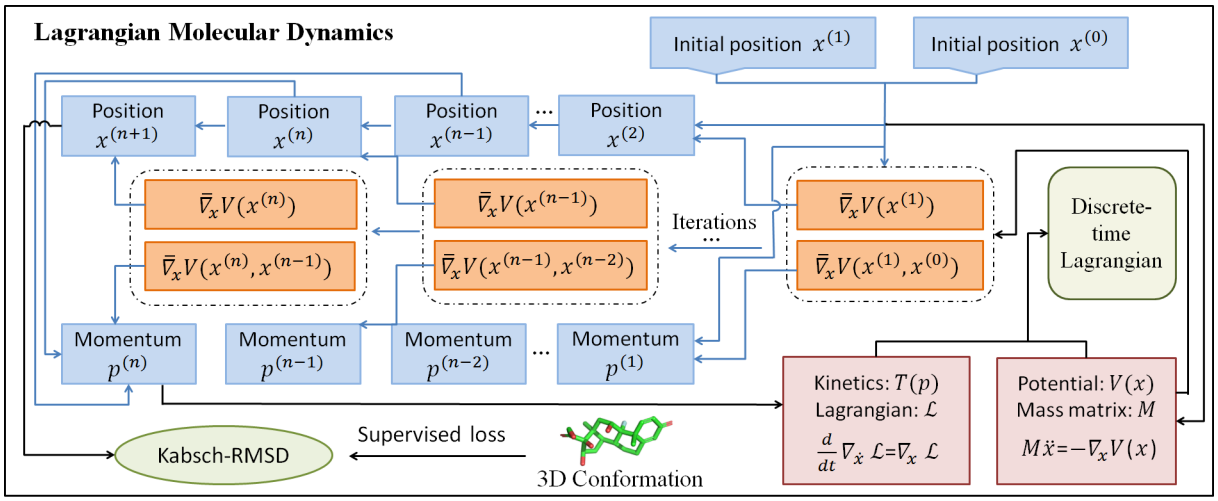


Figure 2: The process of Lagrangian molecular dynamics to simulate molecular force field only with parameterized position, and the final positions predicted by neural network are supervised with the real molecular conformation information.

where z denotes a status $z = (x\dot{x})^\mathsf{T}$ of the system, Δt is the time step size and the n is the n-th time step. Further, the discrete form of Eq. (7) can be expressed as:

$$\frac{p^{(n+1)} - p^{(n)}}{\Delta t} = -\overline{\nabla}_x V(x^{(n+1)}, x^{(n)}) \tag{11}$$

$$\frac{x^{(n+1)} - x^{(n)}}{\Delta t} = \frac{1}{2} M^{-1}(p^{(n+1)} + p^{(n)}) \tag{12}$$

We can obtain $p^{(n+1)}$ from Eq. (11) and Eq. (12) as following:

$$p^{(n+1)} = p^{(n)} - \Delta t \overline{\nabla}_x V(x^{(n+1)}, x^{(n)}) \tag{13}$$

$$p^{(n+1)} = -p^{(n)} + 2M\frac{x^{(n+1)} - x^{(n)}}{\Delta t} \tag{14}$$

Subtract Eq. (14) from Eq. (13) to get $p^{(n)} = M\frac{x^{(n+1)}-x^{(n)}}{\Delta t} + \frac{\Delta t}{2}\overline{\nabla}_x V(x^{(n+1)}, x^{(n)}$, which is substituted into Eq. (13) to get momentum $p^{(n+1)}$ as following:

$$p^{(n+1)} = M\frac{x^{(n+1)} - x^{(n)}}{\Delta t} - \frac{\Delta t}{2}\overline{\nabla}_x V(x^{(n+1)}, x^{(n)}) \tag{15}$$

The momentum $p^{(n)}$ can be expressed from Eq. (15):

$$p^{(n)} = M\frac{x^{(n)} - x^{(n-1)}}{\Delta t} - \frac{\Delta t}{2}\overline{\nabla}_x V(x^{(n)}, x^{(n-1)}) \tag{16}$$

By substituting Eq. (15) and Eq. (16) into Eq. (12), we can obtain $M\frac{x^{(n+1)}-2x^{(n)}+x^{(n-1)}}{(\Delta t)^2} = -\frac{1}{2}(\overline{\nabla}_x V(x^{(n+1)}, x^{(n)}) + \overline{\nabla}_x V(x^{(n)}, x^{(n-1)}))$. Similar to Lagrangian method in physical systems (Aoshima, Matsubara, and Yaguchi 2021), by using $\nabla_x V(x^{(n)})$ to approximate $-\frac{1}{2}(\overline{\nabla}_x V(x^{(n+1)}, x^{(n)}) + \overline{\nabla}_x V(x^{(n)}, x^{(n-1)})$, which is equivalent to the leapfrog integrator (Chen et al. 2020), the variational integrator (Saemundsson et al. 2020) and the Verlet method (Hairer, Wanner, and Lubich 2006), we can obtain the position coordinate $x^{(n+1)}$ at $(n+1)$-th time step as following, that is expressed by the coordinate $x^{(n)}$ at n-th time step and the coordinate $x^{(n-1)}$ at $(n-1)$-th time step:

$$x^{(n+1)} = -\frac{(\Delta t)^2}{M}\nabla_x V(x^{(n)}) - x^{(n-1)} + 2x^{(n)} \tag{17}$$

Obviously, by employing Eq. (15) and Eq. (17), we can calculate the atomic momentum and position of the next time step, which is related to the position and potential energy of the last two steps. In order to accelerate the convergence, the incremental mode is used to iterate in the experiment. Specifically, Figure 2 shows the process of Lagrangian molecular dynamics in detail. Therefore, LagNet is designed only with parameterized position without obeying any additional restrictions, such as the "canonial" constraint, which must be held in Hamiltonian neural networks. LagNet can easily calculate the kinetic energy T by substituting Eq. (15) into Eq. (6). For the potential energy V, we simulate it using simplified Lennard-Jones potential with parameterized distances via a neural network, that can be expressed as:

$$V = \sum_{m \neq n} d_{mn}^{-4} - d_{mn}^{-2} \tag{18}$$

where $d_{mn}^2 = (x_m - x_n)^\mathsf{T} W_V^\mathsf{T} W_V (x_m - x_n)$, W_V is learnable parameter in neural network.

Initializer and loss function. The structure of a molecule can be looked as a graph. Atoms and bonds are regarded as nodes and edges in a graph (Figure 1 (a)), respectively. The original features of atom include atomic symbol, degree, hybridization, the number of radical electrons and so on. The original features of bond are considered as bond types, such as single, double, triple, aromatic, and whether it is a ring. Following the initialization method (Li et al. 2021c), GCN is implemented to extract atoms feature for learning chemical environment in molecule. RNN network is used to determine the initial positions of atoms (Figure 1 (b)), which takes the output of GCN as input. The process of initializer can be expressed as follows:

$$H^{(l+1)} = RNN(GCN(\hat{A}, \hat{D}, H^{(l)})) \tag{19}$$

where GCN and RNN denote graph convolutional network and recurrent neural network, respectively. Let $chunk$ be a partition function, that split $H^{(l+1)}$ into two parts with the same dimension. The initialization can be expressed as: $x^{(0)} = chunk(H^{(l+1)}, 2)[0]$, $x^{(1)} = chunk(H^{(l+1)}, 2)[1]$. In this way, we obtain the initial positions $x^{(0)}$ and $x^{(1)}$.

Following the Kabsch Algorithm (Kabsch 1976) and similar to the method in (Li et al. 2021c), we calculate the Root of Mean Squared Deviations (RMSD) L_{KR} between the predictive atom positions ($\hat{X}$) and the ground-truth conformation positions (X^g) as follows:

$$L_{KR}(\hat{X}, X^g) = (\frac{\sum_{j=1}^{t} m_j \times \parallel \hat{x}_j - x_j^g \parallel_2^2}{\sum_{j=1}^{t} m_j})^{\frac{1}{2}} \tag{20}$$

where m_j denotes atomic mass. In addition, 3-hop distance loss L_{3D} is considered to supervise LagNet training, that can be computed as

$$L_{3D}(\hat{X}, X^g) = (\frac{1}{t} \sum_{j,l=1}^{t} \hat{A}_{jl}^3 (\parallel \hat{x}_j - \hat{x}_l \parallel_2^2 - \parallel x_j^g - x_l^g \parallel_2^2)^2)^{\frac{1}{2}} \tag{21}$$

where $\hat{A}$ denotes the predictive adjacency matrix. Inspired by literature (Li et al. 2021c), we define the final loss function as follows:

$$L_{loss} = L_{KR}(\hat{X}, X^g) + \alpha L_{3D}(\hat{X}, X^g) \tag{22}$$

where α is a predefined hyperparameter.

Experiments

Experiment Settings

Datasets. To evaluate the performance of LagNet with the existing molecular representation learning baselines, we have conducted extensive experiments using different datasets recommended by MoleculeNet (Wu et al. 2018) including a quantum mechanics dataset (QM9), two physiology datasets (Tox21, BBBP) for classification tasks and two physical chemistry datasets (Lipophilicity and FreeSolv) for regression tasks. QM9 (Ramakrishnan et al. 2014) contains geometric conformations (atomic coordinates) and 12 quantitative quantum-chemical properties of 133,885 molecules. Tox21 (Rossoshek 2014) is a dataset about the toxicity of compounds, including toxicity information for 7,831 drug molecules on 12 different targets. BBBP (Martins et al. 2012) contains blood–brain barrier penetration information for over 2,000 compounds on their permeability properties. Lipophilicity (Wenlock and Tomkinson 2015) contains experimental results of octanol/water distribution coefficient (logD at pH 7.4) from 4,200 molecules. FreeSolv (Mobley and Guthrie 2014) provides calculated and experimental hydration free energy of 642 small molecules in water.

Baselines. We compare the proposed LagNet with eleven state-of-the-art models, including 2D-based models (MoleculeNet (Wu et al. 2018), DMPNN (Yang et al. 2019), CMPNN (Song et al. 2020), AttenetiveFP (Xiong et al. 2019), GROVER (Rong et al. 2020b)) and 3D-based models (Drug3D-Net (Li et al. 2021a), GeomGCL (Li et al. 2022b), GEM (Fang et al. 2022), SchNet (Schütt et al. 2017), DimeNet (Klicpera, Groβ, and Günnemann 2020), HamNet (Li et al. 2021c)). MoleculeNet is a widely used benchmark for various methods of molecular representation learning. DMPNN, CMPNN and AttenetiveFP are three variants of message passing neural networks. GROVER is a self-supervised graph transformer method based on molecular motif substructure for learning molecules. Drug3D-Net proposed a spatial-temporal gated attention module to learn molecular voxelized reprersentation. GeomGCL is a geometric graph contrastive learning based on 2D view and 3D view graphs for molecular property prediction. GEM is a geometry-enhanced molecular representation learning method to learn molecular spatial knowledge using 3D structure, such as bond length, bong angles and atomic distances. SchNet models molecular quantum interactions using a continuous-filter convolutional neural network. DimeNet uses the directional information based on angles to learn molecular graph by constructing spherical Bessel function and spherical harmonics. HamNet fits molecular positions and momentums to preserve 3D conformations with a Hamiltonian network.

Implementation Details. LagNet is trained on the basis of Pytorch framework, with Adam [52] as the optimizer to apply gradient back-propagation. All datasets are split into training, validation and testing set with a ratio of 0.8, 0.1, 0.1 , respectively. We use mean absolute error (MAE) for QM9, and root mean squared error (RMSE) for Lipop, FreeSolv to evaluate the performance of regression tasks. As for classification datasets, cross-entropy losses is applied to optimize model, and the receiver operating characteristics curve (ROC) is used as a metric for Tox21 and BBBP to evaluate the overall performance. The learning rate is set to 0.0001 with decay rate 0.00004 during training LagNet. The dropout and batch size are set to 0.2, 16, respectively. It should be noted that the time step size Δt in our experiment is set to 0.025. In addition, we use the identical featurization as HamNet (Li et al. 2021c). In total, 39-dimensional atom features and 10-dimensional bond features are obtained from each molecule.

Performance Evaluation

Overall Comparision. The performance comparisons of conformation prediction on QM9 are presented in Table 1. As shown in Table 1, our model trained in 100 epochs significantly outperforms all the baselines on Kabsch-RMSD loss L_{KR}, 3-hop loss L_{3D} and distance loss based on L_{3D}, which demonstrates the ability of LagNet in reconstructing 3D conformations. On the whole, our proposed LagNet improves the performance over the best HamNet baseline with 6% for Kabsch-RMSD loss. Figure 3 visualizes the conformations at different iterations for molecule "CC(C)C1OCC1=O" selected randomly on test set, (h) depicts the 3D structure of real conformation, (i) is the molecular 2D image generated by RDKit package. As we can see, with the number of iterations increasing, conformations converge to the real one, gradually.

Table 2 summarizes the predictive performances of LagNet and previous 2D-based models and 3D-based mod-

Model	Kabsch-RMSD ($\mathring{A}$)	Distance Loss ($10^{-2}\mathring{A}$)	3-hop Loss ($10^{3}\mathring{A}$)
RDKit	2.349	0.565	4.246
MPNN	1.634	0.592	0.694
HamNet	1.253	0.213	0.411
LagNet (L^v_{loss})	**0.996**	**0.181**	1.742
LagNet (as proposed)	1.180	0.201	**0.402**

Table 1: Performance comparisons of conformation prediction on QM9.

Model Type	Model	Physiology		Physical chemistry			Quantum
		Tox21 Multi-ROC↑	BBBP ROC↑	Lipop RMSE↓	FreeSolv RMSE↓	Pc.Ave RMSE↓	QM9 Multi-MAE↓
2D-based models	MoleculeNet	0.829	0.729	0.655	1.150	0.903	2.350
	DMPNN	0.854	0.917	0.591	1.009	0.800	-
	CMPNN	0.856	0.960	-	0.808	0.808	-
	AttentiveFP	0.858	0.920	0.578	0.768	0.673	1.292
	GROVER	0.831	0.940	0.560	1.544	1.052	-
3D-based models	Drug3D-Net	**0.903**	-	0.993	1.471	1.232	-
	GeomGCL	0.850	-	0.541	0.866	0.704	-
	GEM	0.781	0.724	0.660	1.877	1.269	1.673
	SchNet	0.767	0.847	0.909	3.215	2.062	1.974
	DimeNet	0.780	-	0.614	0.978	0.796	1.920
	HamNet	0.872	-	0.557	0.767	0.662	1.194
	LagNet	0.867	**0.961**	**0.538**	**0.761**	**0.649**	**1.187**

Table 2: Performance comparison of property prediction on various datasets based on baseline models. Baseline results are taken from the ref (Li et al. 2021c, 2022b; Rong et al. 2020b; Fang et al. 2022; Song et al. 2020). "↑" indicates that the larger is better, and "↓" indicates that the smaller is better.

Model	Tox21 (Multi-ROC↑)	BBBP (ROC↑)	Lipop (RMSE↓)	FreeSolv (RMSE↓)	ESOL (RMSE↓)
LagNet (w/o LagM)	0.856	0.952	0.551	0.789	0.558
LagNet (as proposed)	**0.867**	**0.961**	**0.538**	**0.761**	**0.541**

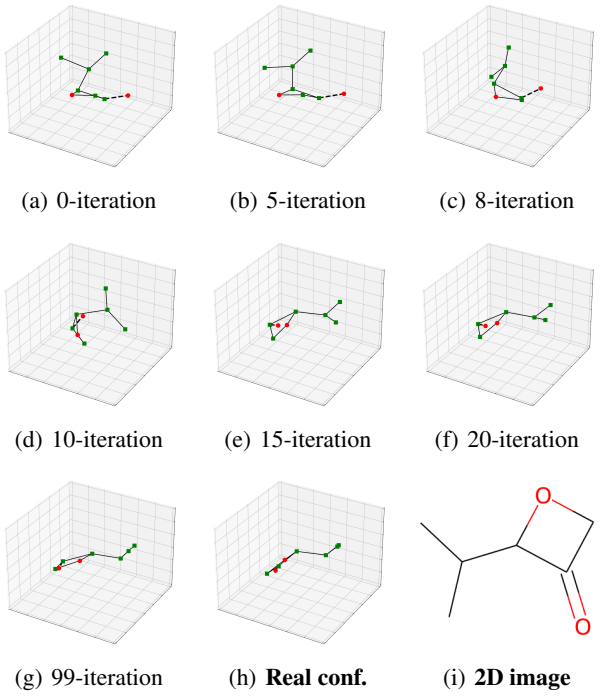
Table 3: Performance comparisons of plug-and-play Lagrangian module on molecular property prediction tasks.

els for the tasks of molecular property prediction with similar network design (Li et al. 2021c). The Pc.Ave denotes the average results of physical chemistry datasets based on RMSE metric. Among all 2D-based models, the well-designed attention mechanics model AttentiveFP and message passing model CMPNN generally show the better performance, which indicates that learning local and nonlocal features in AttentiveFP and the essential bond properties of chemical structure in CMPNN can provide useful semantic information for molecular representation learning. As to 3D-based baseline models, HamNet performs much better than the other baselines due to the integration of interactive implicit positions and momentums based on Hamiltonian network. Although GeomGCL, GEM and DimeNet can identify geometric information such as distance and angle, the 3D information in geometric graphs exists much noise, especially, the 3D position coordinates generation by using Merck Molecular Force Field (MMFF) (Tosco, Stiefl, and Landrum 2014), which is implemented in the RDKit package. Drug3D-Net obtains a high ROC score on Tox21, however, it has poor performance on the other datasets, which may be caused by the boolean expression for voxel-based model. By contrast, LagNet is capable of learning 3D position accurately by simulating molecular force field with discreted Lagrangian neural network. Figure 4 shows the scatter diagrams of 12 quantitative quantum-chemical properties for the test sets on QM9 dataset. As depicted in Figure 4, we observe that the points on the test sets closely surround the identity lines, which illustrates that the predictive results are closer to the target value. Whereas, the property μ of describing dipole moment has a prediction that deviates significantly from the identity line. Thereby, there is a obvious MAE error between the predictive μ and the ground-truth μ.

Ablation Study. To further investigate the loss function that influences the performance of conformation prediction, we conduct the ablation study on Kabsch-RMSD, distance Loss and 3-hop loss with different loss function trained in LagNet. Table 1 shows the performance comparisons with two variants using loss functions L^v_{loss} and L_{loss}, respectively. L_{loss} refers to Eq. (22) and is used to train LagNet, proposed in this paper. LagNet (L^v_{loss}) denotes a variant of LagNet, which is trained with L^v_{loss} loss, including 1-hop loss L_D instead of L_{3D}. L^v_{loss} can be calculated as follows:

$$L^v_{loss} = L_{KR}(\hat{X}, X^g) + \alpha L_D(\hat{X}, X^g) \tag{23}$$

Refer to the previous section for the significance of $\hat{X}$ and X^g. As shown in Table 1, LagNet (L^v_{loss}) leads to better performances on Kabsch-RMSD, distance loss than the pro-

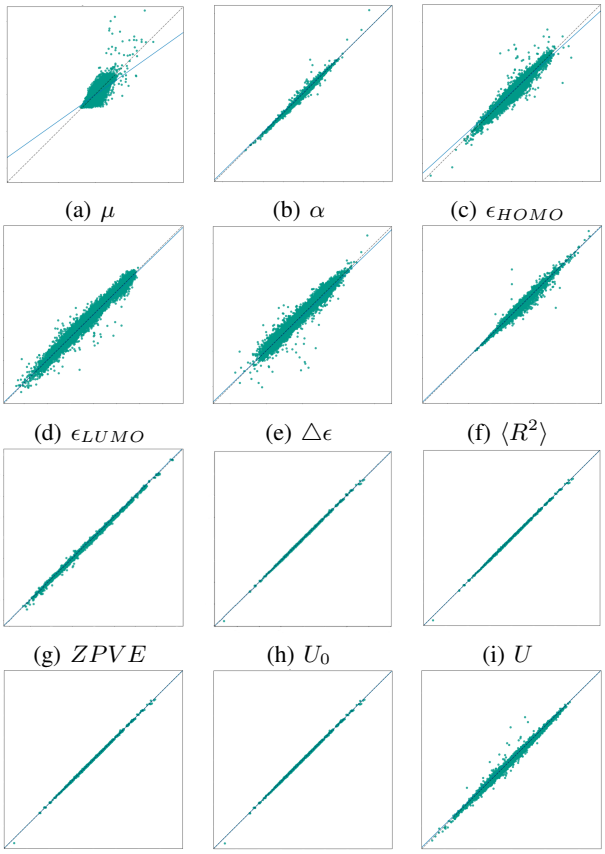


(a) 0-iteration (b) 5-iteration (c) 8-iteration

(d) 10-iteration (e) 15-iteration (f) 20-iteration

(g) 99-iteration (h) **Real conf.** (i) **2D image**

Figure 3: Conformations visualization at different iterations for molecule "CC(C)C1OCC1=O".



(a) μ (b) α (c) ϵ_{HOMO}

(d) ϵ_{LUMO} (e) $\triangle\epsilon$ (f) $\langle R^2 \rangle$

(g) $ZPVE$ (h) U_0 (i) U

(j) H (k) G (l) c_v

Figure 4: Scatter diagrams of 12 quantitative quantum-chemical properties for the test sets on QM9 dataset. The horizontal axis is the predicted value by our model, and the vertical axis is the ground truth value. The trend lines and identity lines for each predicted property are shown as solid lines and the dashed lines, respectively.

posed LagNet. However, the 3-hop loss of LagNet (L^v_{loss}) are obviously large, which indicates although predictive positions tend to appear closer to the labeled locations, the relative structures inside the molecules are distorted and lead to greater reconstruction error for 3D conformation. Whereas the proposed LagNet shows better performance, which is consistent with HamNet (Li et al. 2021c).

Lagrangian Module Analysis. Finally, we analyze the performance variation of plug-and-play Lagrangian module for LagNet with and without molecular force field simulation using discrete Lagrangian mechanics. The performance comparisons are presented in Table 3 with additional dataset ESOL (Delaney 2004). The "w/o LagM" represents without Lagrangian module. As we can see, the proposed model LagNet outperforms LagNet (w/o LagM) on molecular property prediction datasets. The plug-and-play Lagrangian module can simulate molecular force field with parameterized position coordinates using discrete-time Lagrangian equations. Therefore, LagNet improves the predictive performance on various downstream tasks.

Conclusion

In this paper, we propose a novel Lagrangian mechanics-based plug-and-play architecture LagNet by simulating molecular force field for preserving molecular conformation, which brings molecular dynamics simulation into molecular representation learning. LagNet is designed only with parameterized position coordinates, which implements Lagrangian mechanics to learn molecular representation without obeying any additional restrictions. LagNet can generate known conformations and generalize for unknown ones by learning implicit positions iteratively using discrete-time Lagrangian equations. The experimental results show that LagNet can well learn 3D molecular structure features, which thereby demonstrates the effectiveness of the proposed LagNet on molecular property prediction tasks. Meanwhile, LagNet can be also implemented on other tasks in the field of drug discovery, such as drug-drug interaction prediction. Nevertheless, there is still much space for improvement of LagNet. For future work, a straight-forward improvement is to calculate discrete gradient accurately instead of obtaining a approximation using leapfrog integrator. In addition, another promising aspect would be to replace current discrete method with a continuous calculus method, which could enable fine-grained modeling of molecular dynamics. Last but not least, as to noisy data, for providing precise energy and momentum conservation, variational integrator networks (Desai, Mattheakis, and Roberts 2021) will be a considerably interesting topic.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments. This work was supported by the National Natural Science Foundation of China (Grant No. 62262072, U22A2037, 62002304, 62072388), the Collaborative Project fund of Fuzhou-Xiamen-Quanzhou Innovation Zone (No.3502ZCQXT202001), and Fujian Sunshine Charity Foundation.

References

Aoshima, T.; Matsubara, T.; and Yaguchi, T. 2021. Deep discrete-time Lagrangian mechanics. In *International Conference on Learning Representations (ICLR-21) Deep Learning for Simulation (SimDL))*.

Celledoni, E.; Grimm, V.; McLachlan, R. I.; McLaren, D. I.; O'Neale, D.; Owren, B.; and Quispel, G. R. W. 2012. Preserving energy resp. dissipation in numerical PDEs using the "Average Vector Field" method. *Journal of Computational Physics*, 231(20): 6770–6789.

Chen, Z.; Zhang, J.; Arjovsky, M.; and Bottou, L. 2020. Symplectic Recurrent Neural Networks. In *International Conference on Learning Representations (ICLR-20)*, 1–23.

Cho, H.; and Choi, I. S. 2018. Three-Dimensionally Embedded Graph Convolutional Network (3DGCN) for Molecule Interpretation. *arXiv preprint arXiv:1811.09794*.

Cranmer, M.; Greydanus, S.; Hoyer, S.; Battaglia, P.; Spergel, D.; and Ho, S. 2020. Lagrangian neural networks. In *International Conference on Learning Representations (ICLR-20) Deep Differential Equations Workshop*, 1–9.

Delaney, J. S. 2004. ESOL: estimating aqueous solubility directly from molecular structure. *Chemical Information and Computer Sciences*, 44(3): 1000.

Desai, S. A.; Mattheakis, M.; and Roberts, S. J. 2021. Variational integrator graph networks for learning energy-conserving dynamical systems. *Physical Review E*, 104: 035310.

Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; and Wang, H. 2022. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 4: 127–134.

Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural message passing for quantum chemistry. In *International Conference on Machine Learning (ICML-17)*, 1263–1272.

Hairer, E.; Wanner, G.; and Lubich, C. 2006. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer.

Hao, Z.; Lu, C.; Hu, Z.; Wang, H.; Huang, Z.; Liu, Q.; Chen, E.; and Lee, C. 2020. ASGN: an active semi-supervised graph neural network for molecular property prediction. In *Proceeding of the 26th ACM Conference on Knowledge Discovery and Data Mining (KDD-20)*.

Kabsch, W. 1976. A solution of the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32: 922–923.

Kearnes, S.; Mccloskey, K.; Berndl, M.; Pande, V.; and Riley, P. 2016. Molecular graph convolutions: moving beyond fingerprints. *Computer-Aided Molecular Design*, 30(8): 595–608.

Kipf, T.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Klicpera, J.; Groβ, J.; and Günnemann, S. 2020. Directional message passing for molecular graphs. In *International Conference on Learning Representations (ICLR-20)*.

Landrum, G. 2010. RDKit: Open-source cheminformatics.

Li, C.; Feng, J.; Liu, S.; and Yao, J. 2022a. A novel molecular representation learning for molecular property prediction with a multiple SMILES-based augmentation. *Computational Intelligence and Neuroscience*, 2022.

Li, C.; Wang, J.; Niu, Z.; Yao, J.; and Zeng, X. 2021a. A spatial-temporal gated attention module for molecular property prediction based on molecular geometry. *Briefings in Bioinformatics*, 22.

Li, C.; Wei, W.; Li, J.; Yao, J.; Zeng, X.; and Lv, Z. 2021b. 3DMol-Net: Learn 3D molecular representation using adaptive graph convolutional network based on rotation invariance. *IEEE Journal of Biomedical and Health Informatics*.

Li, S.; Zhou, J.; Xu, T.; Dou, D.; and Xiong, H. 2022b. GeomGCL: Geometric Graph Contrastive Learning for Molecular Property Prediction. In *The 36nd AAAI Conference on Artificial Intelligence (AAAI-22)*.

Li, Z.; Yang, S.; Song, G.; and Cai, L. 2021c. HamNet: Conformation-guided molecular representation with hamiltonian neural networks. In *International Conference on Learning Representations (ICLR-21)*.

Lu, D.; Wang, H.; Chen, M.; Liu, J.; Lin, L.; Car, R.; E, W.; Jia, W.; and Zhang, L. 2021. 86 PFLOPS deep potential molecular dynamics simulation of 100 million atoms with ab initio accuracy. *Computer Physics Communicatitons*, 259: 107624.

Luo, S.; Shi, C.; Xu, M.; and Tang, J. 2021. Predicting molecular conformation via dynamic graph score matching. In *International Conference on Neural Information Processing Systems (NeurIPS-21)*.

Lutter, M.; Ritter, C.; and Peters, J. 2019. Deep Lagrangian Networks: Using Physics as Model Prior for Deep Learning. In *International Conference on Learning Representations (ICLR-19)*, 1–17.

Martins, I. F.; Teixeira, A. L.; Pinheiro, L.; and Falcao, A. O. 2012. A Bayesian Approach to in Silico Blood-Brain Barrier Penetration Modeling. *Journal of Chemical Information and Modeling*, 52(6): 1686–1697.

Matsubara, T.; Ishikawa, A.; and Yaguchi, T. 2020. Deep energy-based modeling of discrete-time physics. In *International Conference on Neural Information Processing Systems (NeurIPS-20)*.

Mobley, D. L.; and Guthrie, J. P. 2014. FreeSolv: A database of experimental and calculated hydration free energies, with input files. *Journal of computer-aided molecular design*, 28(7): 711–720.

Paul, A.; Jha, D.; Al-Bahrani, R.; keng Liao, W.; Choudhary, A.; and Agrawal, A. 2018. CheMixNet: mixed DNN architectures for predicting chemical properties using multiple molecular representations. In *Nips Workshop on Machine Learning for Molecules & Materials*.

Ramakrishnan, R.; Dral, P. O.; Rupp, M.; and von Lilienfeld, O. A. 2014. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(140022).

Rogers, D.; and Hahn, M. 2010. Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, 50(5): 742–754.

Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; and Huang, J. 2020a. Self-supervised graph transformer on large-scale molecular data. In *International Conference on Neural Information Processing Systems (NeurIPS-20)*.

Rong, Y.; Bian, Y.; Xu, T.; Xie, W.; Wei, Y.; Huang, W.; and Huang, J. 2020b. Self-Supervised Graph Transformer on Large-Scale Molecular Data. In *International Conference on Neural Information Processing Systems (NeurIPS-20)*, 12559–12571.

Rossoshek, A. 2014. Tox21 Challenge. http://tripod.nih.gov/tox21/challenge/, last accessed on 2020-06-23.

Saemundsson, S.; Terenin, A.; Hofmann, K.; and Deisenroth, M. P. 2020. Variational Integrator Networks for Physically Meaningful Embeddings. In *International Conference on Artificial Intelligence and Statistics*, 1–11.

Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; and Müller, K.-R. 2017. SchNet: a continuous-filter convolutional neural network for modeling quantum interactions. In *International Conference on Neural Information Processing (NIPS-17)*.

Song, Y.; Zheng, S.; Niu, Z.; Fu, Z.-H.; Lu, Y.; and Yang, Y. 2020. Communicative Representation Learning on Attributed Molecular Graphs. In *International Joint Conference on Artificial Intetlligence (IJCAI-20)*.

Tosco, P.; Stiefl, N.; and Landrum, G. 2014. Bringing the MMFF force field to the RDKit: implementation and validation. *Journal of Cheminformatics*, 6(37): 1–4.

Wand, W.; Xu, M.; Cai, C.; Miller, B. K.; Smidt, T.; Wang, Y.; Tang, J.; and Gómez-Bombarelli, R. 2022. Generative coarse-graining of molecular conformations. In *International Conference on Machine Learning (ICML-22)*.

Wenlock, M.; and Tomkinson, N. 2015. Experimental in vitro DMPK and physicochemical data on a set of publicly disclosed compounds. https://doi.org/10.6019/chembl3301361, last accessed on 2020-10-29.

Wu, F.; Zhang, Q.; Radev, D.; Cui, J.; Zhang, W.; Xing, H.; Zhang, N.; and Chen, H. 2021. Molformer: motif-based Transformer on 3D heterogeneous molecular graphs. *arXiv preprint arXiv:2110.01191*.

Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; and Pande, V. S. 2018. MoleculeNet: A Benchmark for Molecular Machine Learning. *Chemical Science*, 9(2): 513–530.

Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; and Zheng, M. 2019. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. *Journal of Medicinal Chemistry*.

Xu, M.; Luo, S.; Bengio, Y.; Peng, J.; and Tang, J. 2021a. Learning gradient fields for molecular conformation generation. *arXiv preprint arXiv:2105.03902*.

Xu, M.; Luo, S.; Bengio, Y.; Peng, J.; and Tang, J. 2021b. Learning neural generative dynamics for molecular conformation generation. *arXiv preprint arXiv:2102.10240*.

Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; and Tang, J. 2022. GeoDiff: a geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations (ICLR-22)*.

Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; and Barzilay, R. 2019. Are learned molecular representations ready for prime time? *arXiv preprint arXiv:1904.01561*.

Yang, S.; Li, Z.; Song, G.; and Cai, L. 2021. Deep molecular representation learning via fusing physical and chemical information. In *International Conference on Neural Information Processing Systems (NeurIPS-21)*.

Zhang, L.; Han, J.; Wang, H.; Car, R.; and E, W. 2018. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics. *Physical review letters*, 120: 143001.

Zhong, Y. D.; Dey, B.; and Chakraborty, A. 2020. Symplectic ODE-Net: Learning Hamiltonian Dynamics with Control. In *International Conference on Learning Representations (ICLR-20)*, 1–17.